RADC-TDR-62-567

Final Technical Report

OPTIMUM SPEECH SIGNAL MAPPING TECHNIQUES

TECHNICAL DOCUMENTARY REPORT NO. RADC-TDR-62-567

28 September 1962

Electronic Warfare Laboratory
Rome Air Development Center
Research and Technology Division
Air Force Systems Command
United States Air Force
Griffiss Air Force Base, New York

Program Element Code Nr. - 62405454
760D, Project 4027, Task 402704
Electronic Systems Division (RADC)
Mr. Chapin, RAWER

(Prepared under Contract No. AF30(602)-2715 by Sylvania Electronic Systems, 40 Sylvan Road, Waltham 54, Mass.)

FOREWORD

This report embodies results of a continued program of research to implement and test methods of mathematically mapping speech signals, employing the theory and methods of orthogonalized exponentially damped sinusoidal functions. The objective of this program was to obtain a representation for speech signals that is optimal with respect to preservation of information content, simplicity in implementation and information efficiency.

AIR FORCE, Cato Show Prtg. Co. 2/11/63 - 69

ABSTRACT

Investigations of the analysis of speech in terms of a fixed exponential function series have been carried out. The analysis-synthesis processing was performed via digital computers operating with digitized speech. The results indicate that the representation used is an efficient one for speech waveform analysis and that the information content of the speech is preserved when phase information is eliminated. The spectral coefficients after phase elimination are not found to be an efficient representation for the amplitude spectrum. Other results show that the method of analysis is not limited to the speech of one individual. Analytical studies indicate that it is possible to optimize the method of analysis to essentially perfect its efficiency for speech waveform analysis. Normalized, phase eliminated, spectral patterns derived for ten vowel utterances by five talkers indicate the feasibility of performing both automatic vowel and/or automatic speaker recognition using the orthonormal coefficient data.

PUBLICATION REVIEW

This report has been reviewed and is approved.

Approved: Edward N. Munzer
EDWARD N. MUNZER
Chief, Electronic Warfare Laboratory
Directorate of Intelligence & Electronic Warfare

Approved: Robert J. Quinn, Jr.
ROBERT J. QUINN, JR., Col, USAF
Director of Intelligence & Electronic Warfare

FOR THE COMMANDER: Irving J. Gabelman
IRVING J. GABELMAN
Director of Advanced Studies

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

LIST OF ILLUSTRATIONS

SECTION 1

INTRODUCTION

The Optimum Speech Signal Mapping Techniques program has continued the study and evaluation of speech waveform analysis methods initiated under a previous contract, No. AF30(602)-2446 with Rome Air Development Center. Under that program, analysis-synthesis of continuous speech in terms of a fixed set of orthogonalized exponentially damped sinusoidal functions was successfully demonstrated. The program, herein reported, continues and extends the work of the previous contract through the completion of the following items:

(1) Tests were made, of the correlations among the orthonormal coefficients of the fixed exponential set for a six second sample of speech, using the orthonormal coefficients data available from the previous contract. The low cross correlations observed indicate that the function series used is an efficient one for speech waveforms.

(2) Tests were also made of the cross correlations among the coefficients after phase information is eliminated. This refers to a set of coefficients representing the spectrum magnitude. The correlations in this case were substantially higher, indicating that the function series used does not yield a particularly efficient representation of the amplitude spectrum.

(3) A preliminary study of the power spectrum of one set of coefficients was made to determine the feasibility of time averaging the coefficients. The results indicated that it is feasible to time average the coefficients since most of their energy spectrum lies below 20 cycles/second.

(4) A test of the ability of the orthogonalized exponentially damped sinusoids to accurately represent the input speech when the phase data have been eliminated from the representation was conducted. The results show that the speech is still intelligible, indicating that the phase elimination operation can be used to reduce the data by a factor of two for speech recognition purposes. The discontinuities at the pitch period boundaries make the quality of the resynthesized phase eliminated speech objectionable.

(5) Analytical studies, based in part upon the Karhunen-Loeve expansion were conducted in an effort to find optimum functional forms for speech analysis. The results indicate that several methods exist for deriving a new function set wherein each new function is a weighted sum of the basic exponentially damped sinusoids. The new set would be derived using experimental data on coefficient cross-correlations and would be a more nearly optimum representation for speech. However, results to date indicate that such new sets probably will be only slightly greater in efficiency than the basic set.

(6) Individual pitch periods taken from the central portions of ten vowels, as uttered by five different male talkers, were analyzed and resynthesized using orthogonalized exponentially damped sinusoids. The results indicate that the analysis-synthesis method operates satisfactorily on the speech of different talkers.

(7) The spectral information for the ten vowels, obtained from the above analysis, has been normalized and plotted to indicate the different patterns exhibited for each vowel. While both more quantitative analysis and more data would be needed for a final proof, it seems quite likely that enough information is contained in these patterns to differentiate the vowels in an automatic speech recognition program. Speaker to speaker variations in these patterns indicate that the orthonormal coefficient data can be made the basis of automatic speaker recognition particularly in view of the fact that the speakers all were of New England background.

The discussion which follows contains the details of the methods used in accomplishing the tasks proposed on this program together with those results which can be visually presented.

SECTION 2

DISCUSSION

2.0 PRELIMINARIES

It is worthwhile to present the following brief review of the basic analysis-synthesis methods being used in this program in order to define the terms used later.

The speech analysis technique being studied under this program operates by sectioning the waveform into "analysis intervals." These intervals correspond to pitch periods for the voiced sounds. If $S(t)$ is the speech waveform, then the waveform of the q^{th} analysis interval is

$$S_q(t) = S(t),\ t_q \leq t \leq t_{q+1} \tag{1}$$

$$= 0$$

so that

$$S(t) = \sum_q S_q(t) \quad . \tag{2}$$

During the q^{th} analysis interval, the waveform is represented by an approximation of the form:

$$S_q(t) \approx \sum_{K=1}^{n/2} A_k^q \, e^{-\alpha_k(t-t_q)} \sin\left[\beta_k(t - t_q) + \eta_k^q\right] \tag{3}$$

where

$$n = 16,\ \beta_k = 2\pi k\ 200,\ K = 1,2,\cdots,16 \ .$$

$$\alpha_k = \frac{\beta_k}{20} \tag{4}$$

Since the A_k^q and η_k^q are difficult to compute for a minimum mean square error approximation, the analysis-synthesis process is actually performed in terms of an orthogonalized version of the exponentially damped functions in Eq. (3). Denote this set of orthonormal functions by $\left\{\phi_u(t)\right\}$. We then write

$$S_q(t) \approx \sum_{u=1}^{n} C_u^q \, \phi_u(t - t_q) \tag{5}$$

where

$$C_u^q = \int_{t_q}^{t_{q+1}} S(t) \, \phi_u(t - t_q) \, dt \tag{6}$$

and the $\left\{\phi_u(t)\right\}$ have the orthonormal property over $0 \leq t \leq \infty$:

$$\int_0^\infty \phi_u(t) \, \phi_\nu(t) \, dt = 1 \qquad u = \nu$$
$$= 0 \qquad u \neq \nu \tag{7}$$

The C_u^q are referred to as the orthonormal coefficients of the waveform $S_q(t)$. Details of the orthogonalization process are given in Reference 1. The orthogonalized functions $\left\{\phi_u(t)\right\}$ occur in pairs. Both members of such a pair have the same amplitude spectrum, i.e., they differ only in their phase spectra. The series in Eq. (5) can be rewritten to place these pairs in evidence as follows:

$$S_q(t) \approx \sum_{\nu=1}^{n/2} \left[C_{2\nu-1}^q \, \phi_{2\nu-1}(t - t_q) + C_{2\nu}^q \, \phi_{2\nu}(t - t_q)\right] \tag{8}$$

A "phase elimination" operation that is analogous to the same operation in a Fourier sine-cosine series can be carried out giving:

$$S'_q(t) = \sum_{\nu=1}^{n/2} B^q_\nu \phi_{2\nu}(t - t_q) , \tag{9}$$

where

$$B^q_\nu = \left[(C^q_{2\nu-1})^2 + (C^q_{2\nu})^2\right]^{1/2} . \tag{10}$$

The B^q_ν are referred to as the phase eliminated orthonormal coefficients of the waveform $S_q(t)$.

The functions used in Eq. (9) could as well have been the odd ordered functions $\phi_{2\nu-1}(t)$.

Another phase eliminated resynthesis which yields smaller discontinuities at the times $t = t_q$ (pitch period boundaries) in the resynthesized speech is given by:

$$S''_q(t) = \sum_{\nu=1}^{n/2} (-1)^{\nu-1} B^q_\nu \phi_{2\nu}(t - t_q) \tag{11}$$

The details of the methods by which digital computers are used to process the speech to obtain $S_q(t)$, $S'_q(t)$ and $S''_q(t)$ are discussed in Reference 1 and in Appendix III.

2.1 ANALYSIS-SYNTHESIS

At the outset of the six month research program, which is the subject of this report, a six second sample of connected speech comprising the two sentences "Joe took father's shoe bench out. She is waiting at my lawn.", had already been analyzed via Eq. (6) and resynthesized via Eqs. (5) and (2), (see Reference 1). This demonstrated the validity of the particular series representation described in Reference 1. During the current program, the same speech sample was also resynthesized using the same coefficient data derived by Eq. (6), but with the "phase" information eliminated via Eq. (10).

This operation was performed to show that, by phase elimination, the number of coefficients used to represent a given sound could be cut in half. This is based upon the facts that the odd and even pairs $\phi_{2\nu-1}^{(t)}$ and $\phi_{2\nu}^{(t)}$ differ only in their phase spectra and that the ear is relatively insensitive to changes in the phase spectrum of a quasi-periodic wave. Both the resynthesis using Eq. (9) for n = 32, 24, and 16, and that using Eq. (11) for n = 32 were performed. Since all of the orthonormal functions, in the set being used, have positive values at t = 0, and since all of the coefficients B_{ν}^{q} are positive, a positive initial value for each pitch period resynthesized via Eq. (9) results. The alternating sign resynthesis of Eq. (11) reduces these discontinuities. An example of the result for a single pitch period is shown in Figure 1. Note that the resynthesis via Eq. (11) reduces but does not eliminate the discontinuity at the start of the pitch period.

A listening test indicates that the phase eliminated speech is about as intelligible as the non-phase eliminated speech, but that the quality of the speech is objectionable because of the sharp discontinuities between pitch periods.

This effect is less pronounced but is still objectionable in the resynthesis using alternating coefficient signs. (See Eq. (11).)

In order to show that the analysis-synthesis operations described above preserve the formant patterns in the speech sample, wide band sound spectrograms were made of 2.4 second portions of the processed speech. The spectrograms are made from the words "Joe took father's shoe ben...." Figure 2a shows the spectrogram for the speech sample that was low pass filtered to 3000 cycles per second bandwidth, and subjected to analog to digital and digital to analog conversion. Figure 2b shows a spectrogram for the resynthesized speech using the 32 function orthonormal expansion. The spectrogram in Figure 2c was made from the phase eliminated resynthesized speech sample using Eq. (11). The formant pattern of the original speech is seen to have been recreated in the resynthesized versions. The discontinuities between pitch periods tend to "fill in" areas that are blank in the original speech. This is especially true in Figure 2c where the phase elimination process causes greater discontinuities to occur.

2.2 EFFICIENCY OF THE ORTHONORMAL SERIES REPRESENTATION

It is of interest to determine whether the orthonormal series that we are using yields the most accurate possible description of the speech signal for a given number of terms in the series. Given, that the orthonormal coefficients have been shown to yield a complete description of the speech, one may use the correlation coefficients

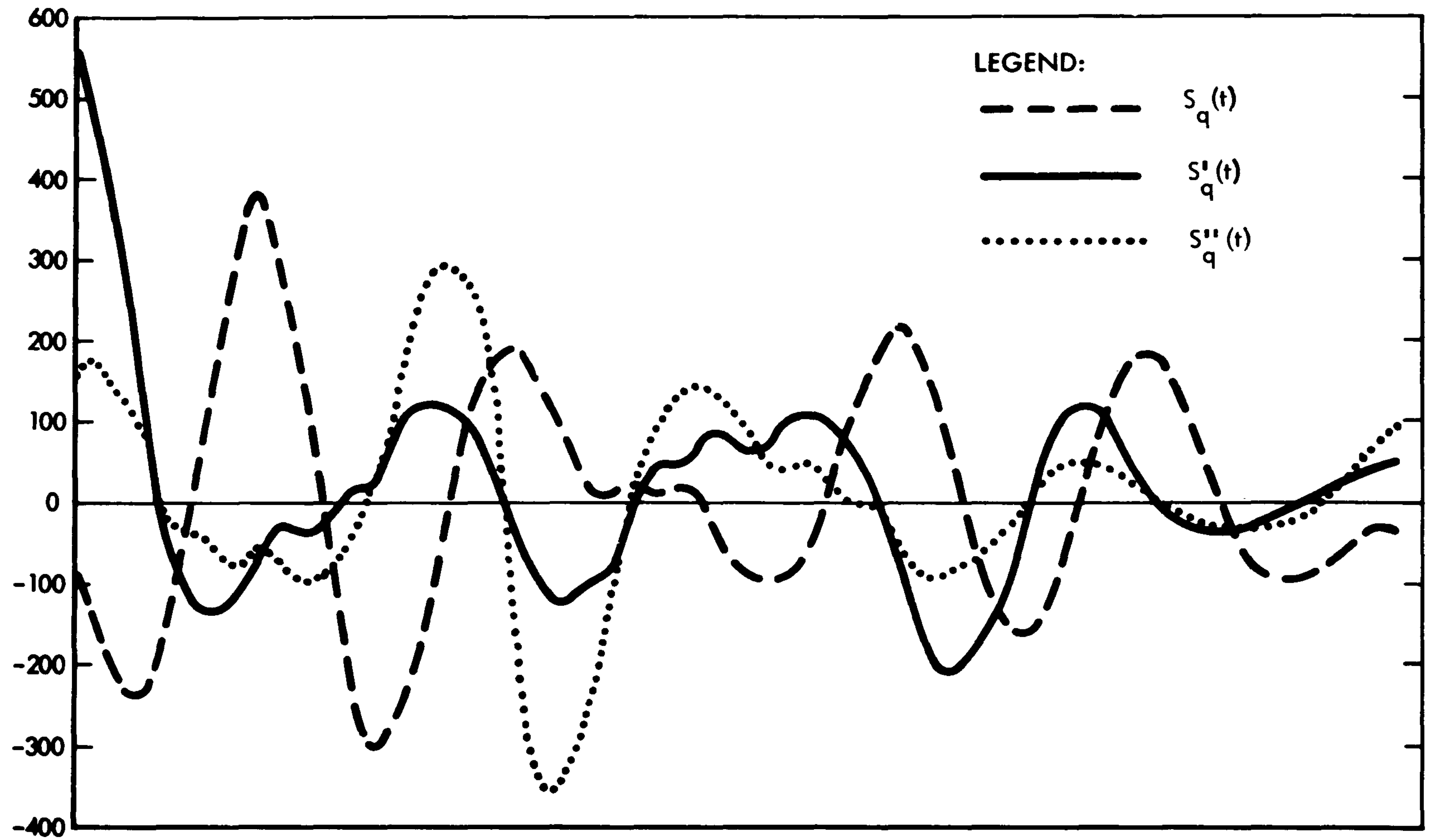


Figure 1. Phase Eliminated Speech Resynthesis

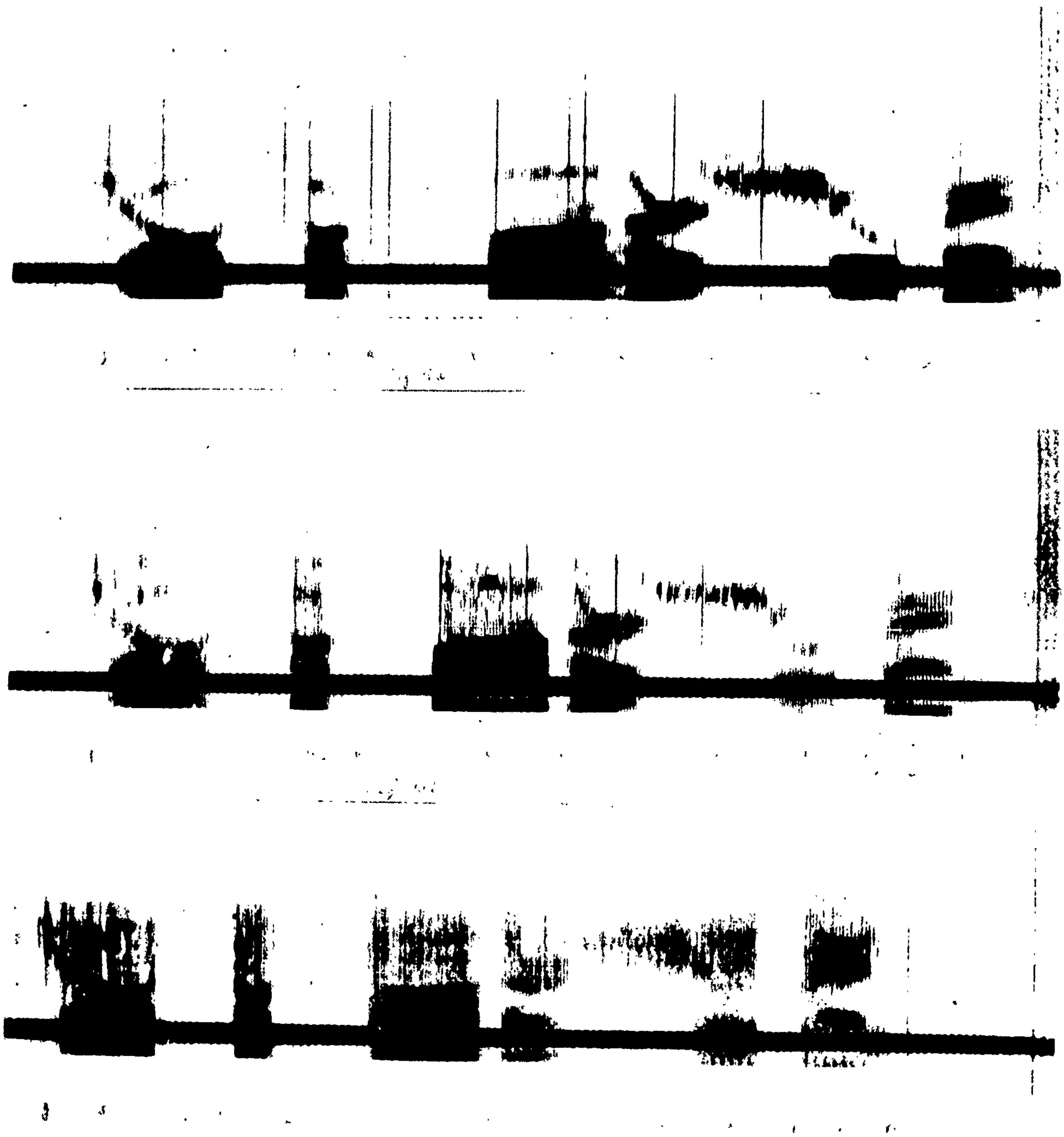


Figure 2. Sound Spectrograms for Original Speech, Resynthesized Speech, and Phase Eliminated Resynthesized Speech

ρ_{uv} among the various coefficients as a measure of whether or not each coefficient yields an independent measurement on the speech signal. The most desirable situation is given by:

$$\begin{aligned} \rho_{uv} &= 1 \qquad u = v \\ &= 0 \qquad u \neq v \end{aligned} \tag{12}$$

It is, of course, true that the linear independence implied by Eq. (12) does not, in general, prove statistical independence, but under the circumstances this is the most reliable indication that is readily available. Accordingly, the data available from the previous analysis of six seconds of connected speech (which included nearly all of the commonly occurring phonemes of english) were used to compute

$$\rho_{uv} = \frac{E\left[(C_u - \bar{C}_u)(C_v - \bar{C}_v)\right]}{\sigma_u \sigma_v} , \tag{13}$$

where E [ ] indicates the expected value, the over bar indicates the mean value, and

$$\sigma_u = \left[E(C_u - \bar{C}_u)^2\right]^{1/2} , \quad \text{and}$$

$$u = 1,2,3,\cdots, 32$$

$$v = 1,2,3,\cdots, 32 \quad .$$

All averages are over the total number of pitch periods. The matrix whose elements are the ρ_{uv} is given in Figure 3. The cross-correlations indicate that the representation is an efficient one for speech, since only about 26 percent of the correlation coefficients are greater than 0.25, and only 1.6 percent of the correlation coefficients are 0.50 or greater.

The same type of cross-correlation analysis was conducted using the phase eliminated orthonormal coefficients B_ν (see Eq. 10). In this case the object is to determine whether the B_ν represent essentially independent measurements on the magnitude spectrum of the speech. The cross-correlation coefficients are given by

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	
1	.24	.18	-.49	-.10	-.27	-.23	.17	-.00	.04	.21	-.05	-.08	-.14	.12	-.22	-.01	-.13	.08	-.15	.11	-.10	.12	-.21	.14	-.30	.22	-.27	.19	-.31	.19	-.30	1
	1	.38	-.26	-.04	-.29	-.10	-.04	.06	.07	.09	-.16	.09	-.18	.15	-.07	.19	-.19	.12	-.15	.13	-.09	.13	-.11	.17	-.16	.24	-.10	.23	-.11	.30	-.13	2
		1	.11	-.22	-.31	-.28	.01	-.01	.14	-.01	-.07	-.10	-.10	.06	-.12	.01	-.22	.03	-.01	.08	-.10	.04	-.15	.10	-.15	.16	-.17	.16	-.14	.20	-.14	3
			1	.22	.09	-.03	-.33	-.19	-.10	-.09	-.02	-.05	-.01	-.04	.12	.05	-.04	.004	-.02	-.03	.01	.01	.001	.04	.10	.03	.04	.03	.13	-.02	.13	4
				1	.29	-.20	-.40	-.57	-.04	.12	.15	.05	.08	-.07	.09	.01	.01	-.05	.01	-.04	.05	.02	.06	-.01	.01	-.01	-.01	-.03	-.04	-.0001	.11	5
					1	.37	-.49	-.24	-.47	-.07	.09	-.0001	.01	-.02	-.11	-.07	.10	.02	.02	.08	-.07	.04	.09	.04	-.003	.02	-.04	.07	.01	.07	.07	6
						1	-.27	.33	-.41	-.24	.004	.11	-.002	-.04	.07	.11	.09	.09	.07	.09	.07	.03	.09	.02	.14	-.01	.24	.05	.30	.04	.26	7
							1	.21	.45	-.15	.07	-.27	.10	-.25	.02	.22	.11	-.20	.08	-.24	.13	-.23	.07	-.23	.06	-.31	.07	-.36	.02	-.31	-.03	8
								1	-.08	-.28	.02	-.02	.09	-.06	.14	.06	.10	.04	.07	-.001	.04	-.03	.11	-.09	.20	-.09	.29	-.05	.26	-.12	.21	9
									1	-.30	-.14	-.31	.03	-.26	.001	-.18	.06	-.16	.03	-.19	.04	-.22	-.003	-.18	.08	-.27	.03	-.26	.002	-.21	-.09	10
										1	.07	.04	-.04	.01	-.05	.06	-.05	.09	.002	.14	-.04	.10	-.07	.10	-.11	.10	-.10	.09	-.09	.08	-.04	11
											1	-.20	.33	-.24	.09	-.33	.24	-.31	.24	-.17	.22	-.24	.20	-.36	.14	-.35	.23	-.41	.26	-.44	.25	12
												1	-.19	.34	-.35	.09	.06	.16	.06	.26	.05	.14	-.09	.12	-.002	.23	-.001	.14	-.06	.17	.06	13
													1	-.01	.36	-.31	.15	-.41	.28	-.29	.24	-.25	.26	-.39	.29	-.43	.33	-.47	.34	-.50	.30	14
														1	-.05	-.01	-.37	-.01	-.17	.16	-.03	.26	-.11	.16	-.21	.32	-.22	.31	-.23	.33	-.16	15
															1	.16	-.21	-.19	-.10	-.34	.18	-.07	.14	-.09	.19	-.30	.20	-.13	.33	-.17	.20	16
																1	-.27	.05	-.27	-.08	.01	.32	-.07	.19	-.08	.18	-.17	.35	-.01	.36	-.07	17
																	1	.11	.21	-.12	-.15	-.37	.32	-.35	.29	-.30	.33	-.30	.24	-.38	.20	18
																		1	-.12	.13	-.52	-.20	-.06	.36	.02	.39	-.02	.34	-.26	.24	.04	19
																			1	.15	.07	-.38	.02	-.41	.39	-.36	.34	-.32	.27	-.36	.21	20
																				1	-.13	.02	-.44	.06	.02	.27	-.02	.28	.01	.25	-.21	21
																					1	.16	-.05	-.30	-.03	-.34	.16	-.38	.29	-.15	.24	22
																						1	-.17	.12	-.42	.30	-.23	.27	-.17	.36	-.05	23
																							1	-.14	.06	-.36	.21	-.33	.25	-.35	.33	24
																								1	-.25	.48	-.47	.33	-.27	.41	-.12	25
																									1	-.32	.46	-.31	.40	-.51	.39	26
																										1	-.35	.52	-.53	.48	-.21	27
																											1	-.37	.56	-.49	.37	28
																												1	-.39	.60	-.44	29
																													1	-.32	.27	30
																														1	-.42	31
																															1	32

Figure 3. Correlation Matrix of Orthonormal Coefficients

$$\Gamma_{\mu\nu} = \frac{E\left[(B_\nu - \bar{B}_\nu)(B_\mu - \bar{B}_\mu)\right]}{\sigma_\mu \sigma_\nu} , \tag{14}$$

where

$$\sigma_\mu = \left[E(B_\mu - \bar{B}_\mu)^2\right]^{1/2}$$

and

$$\mu = 1,2,\cdots,16, \quad \nu = 1,2,\cdots,16 .$$

The matrix of correlation coefficients is shown in Figure 4. The cross-correlation coefficients are generally much higher than those of Figure 3, indicating that measurement of the B_ν is not a particularly efficient method of describing the amplitude spectrum.

2.3 APPROXIMATING OPTIMUM ORTHONORMAL FUNCTIONS FOR SPEECH ANALYSIS

One of the goals of the present program is the optimization of a set of orthonormal functions for the purpose of speech analysis. The desired properties are (a) minimum mean square truncation error, and (b) linearly independent coefficients. The Karhunen-Loeve theorem (see Reference 1 and 2 of Appendix II) indicate that these properties are provided by the eigenfunctions of the homogeneous Fredholm equation whose kernel is the covariance characteristic of the speech process. It is unlikely that an exact solution of this problem will be obtained. This is because it appears to be impossible to define an experiment that would yield meaningful data for the specification of a covariance function for speech. The reader is referred to Appendix II for a more extensive discussion of these difficulties, as well as detailed mathematical treatment of the work summarized below.

As a result of the theoretical work carried out on this problem, it is possible to define two methods of solution of the problem of obtaining an optimal function set. Both methods start from a basic Fourier type analysis using an arbitrary function set. (In the speech waveform analysis case, the basic function set could be the orthogonalized exponentially damped sinusoids, since they have already been shown to be efficient for speech and also have the convenient property of being able to handle

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1	.6	.19	.10	.15	.26	.32	.24	.24	.27	.16	.16	.29	.34	.36	.40	1
	1	.32	.22	.22	.37	.51	.44	.44	.45	.30	.33	.38	.42	.45	.49	2
		1	.69	.85	.61	.48	.46	.39	.32	.28	.30	.51	.47	.53	.53	3
			1	.80	.71	.51	.44	.28	.24	.25	.28	.34	.42	.48	.48	4
				1	.69	.48	.37	.27	.26	.25	.25	.47	.49	.53	.53	5
					1	.72	.51	.34	.34	.32	.30	.45	.47	.50	.55	6
						1	.71	.46	.46	.42	.38	.47	.54	.58	.61	7
							1	.73	.63	.59	.60	.54	.59	.61	.65	8
								1	.68	.60	.62	.48	.47	.51	.54	9
									1	.68	.69	.64	.63	.60	.64	10
										1	.69	.52	.56	.60	.60	11
											1	.63	.62	.63	.63	12
												1	.81	.71	.76	13
													1	.88	.84	14
														1	.83	15
															1	16

Figure 4. Correlation Matrix of Phase Eliminated Orthonormal Coefficients

arbitrary length pitch periods.) Using the experimentally measured cross-correlations between the orthonormal coefficients (essentially an un-normalized version of the data given in Figure 2), a new set of functions is computed. Each of the new functions is an explicit linear summation of the old functions wherein the weights are computed (using the data of Figure 2), such that the statistical average of the cross-correlations between coefficients of the optimized functions, has been minimized (see Appendix II, Eq. (17)).

The optimized functions may be constrained, if desired, to be orthonormal (Appendix II) in which case the resulting solutions are equivalent to solutions of the Karhunen-Loeve equation (Appendix II).

On the other hand, the requirement for orthonormality can be relaxed. In this case, explicit expressions for the optimum weights are not obtained, but instead, the problem of minimization of a matrix involving the cross-correlations between orthonormal coefficients (of the basic series) must be solved. This latter method, although not explicit, may be capable of producing a function set that will yield a truncation error as low or lower than the aforementioned explicit method.

Using the theory developed in Appendix II, and the available experimental data described in Section 2.2 of this report, it will be possible to determine how much of an improvement in efficiency over the basic orthogonalized exponentially damped sinusoids can be obtained through the use of optimum linear weighting.

2.4 ENERGY SPECTRUM OF THE ORTHONORMAL COEFFICIENTS

To the extent that the orthonormal coefficients are a spectral description of the speech signal, it may be expected that the rate of variation of the quantities $\left\{B_{\nu}^{q}\right\}$ will be limited by the rate of variation of the spectrum controlling mechanism -- the vocal tract. The results of a hand-computed, octave band, spectral analysis of a 0.45 second record of B_{3}^{q} are shown in Figure 5. Figure 5 shows that, for the particular phase eliminated coefficient considered, there is relatively little energy at rates above 18 cycles per second. This value is typical of the parameter bandwidths that have been found necessary in speech analysis-synthesis systems such as the channel Vocoder.

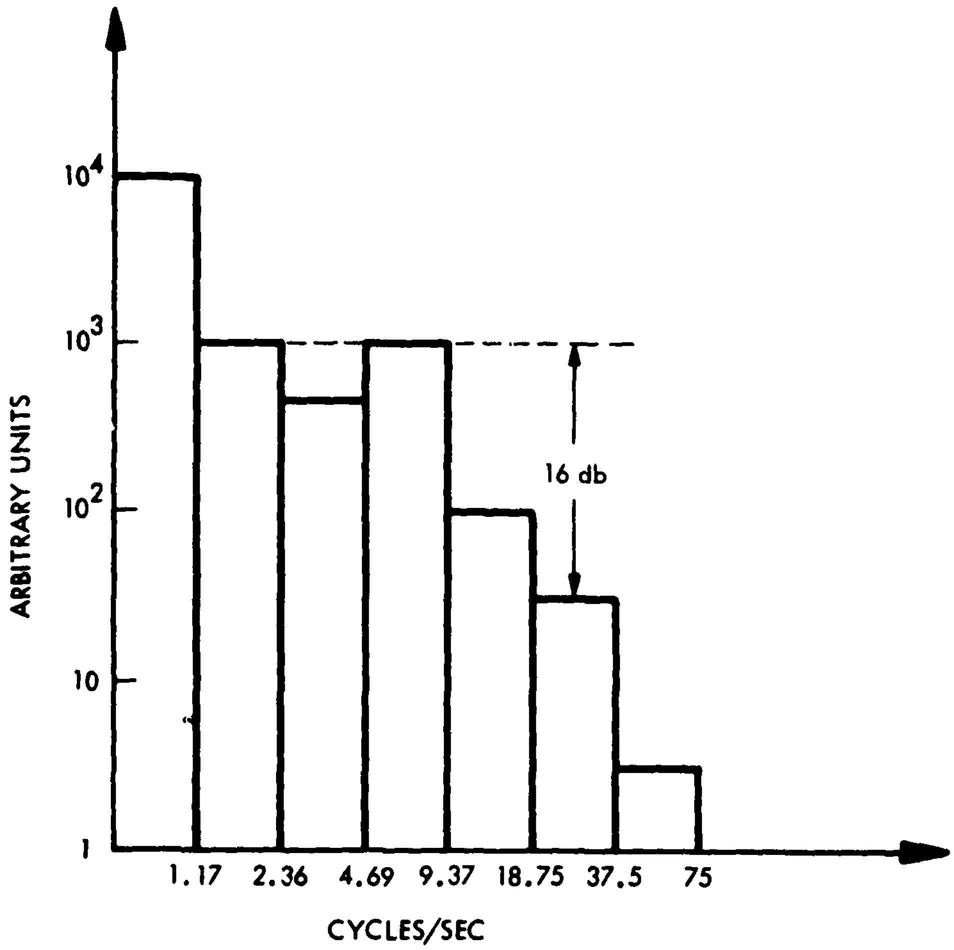


Figure 5. Energy Spectrum of Typical Phase Eliminated Orthonormal Coefficient

2.5 TESTS OF THE ANALYSIS METHOD ON SPEECH SAMPLES TAKEN FROM SEVERAL TALKERS

One of the limitations of the previous tests of speech analysis-synthesis using orthogonalized exponentially damped sinusoids has been removed, during the current program, by making tests on the speech of five male talkers. Two samples of each of the ten words: he, hid, head, had, hub, her, ah, awe, hood, who, were recorded from each of five speakers and digitized at a sampling rate of 12,000 samples/second with 10 bits per sample accuracy. Plots were made of the waveforms of these words so that individual pitch periods could be selected for analysis. A pitch period from the central portion of the vowel in each of the 100 digitized words was selected for analysis. These 100 pitch periods were then analyzed using Eq. (6), the full 32 term orthonormal function series. The orthonormal coefficients were saved for later analysis. These coefficients and the function set were used to resynthesize least mean square error approximations to the original waveforms. Fifty of these pitch period curves (one for each of ten vowels for each of five speakers) are shown in Appendix I. These provide ample opportunity for comparison of original and resynthesized curves, and show that the particular choice of functions and parameters used in the analysis work of this program are not limited in applicability to any particular speaker.

2.6 PATTERNS OF PHASE ELIMINATED ORTHONORMAL COEFFICIENTS FOR TEN VOWELS

The orthonormal coefficient data for the five speakers (as described in Section 2.4) were used to compute the phase eliminated orthonormal coefficients as given by Eq. (10) of Section 2.0. The coefficients for each of the two repetitions by each speaker of each word were normalized and then averaged. These normalized and averaged coefficients are plotted in Figures 6 through 15 for the ten words as described above in Section 2.4. Each of these figures contains the coefficient pattern for one pitch period of the central vowel of the given word for each of the five talkers. Examination of these patterns shows that different patterns are indeed obtained for the different vowels. Thus, the data indicate that such coefficient patterns would be useful input data for a speech recognition machine.

It should be noted that the results shown are for single pitch periods, and that in a practical situation more reliable results would probably be obtained by using data obtained from several adjacent pitch periods in each sound to be recognized.

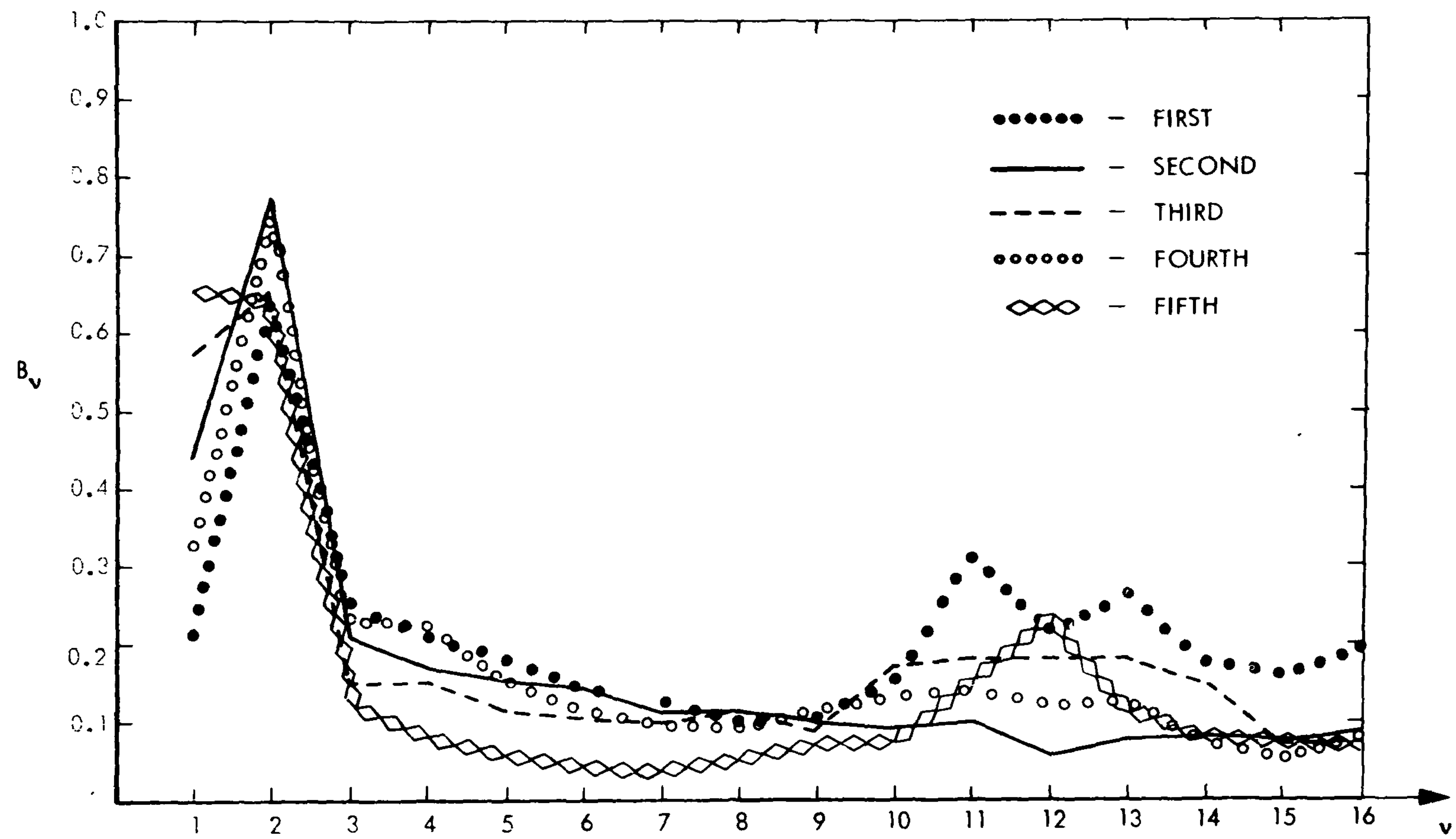


Figure 6. Orthonormal Spectral Pattern for h<u>e</u> (Vowel)

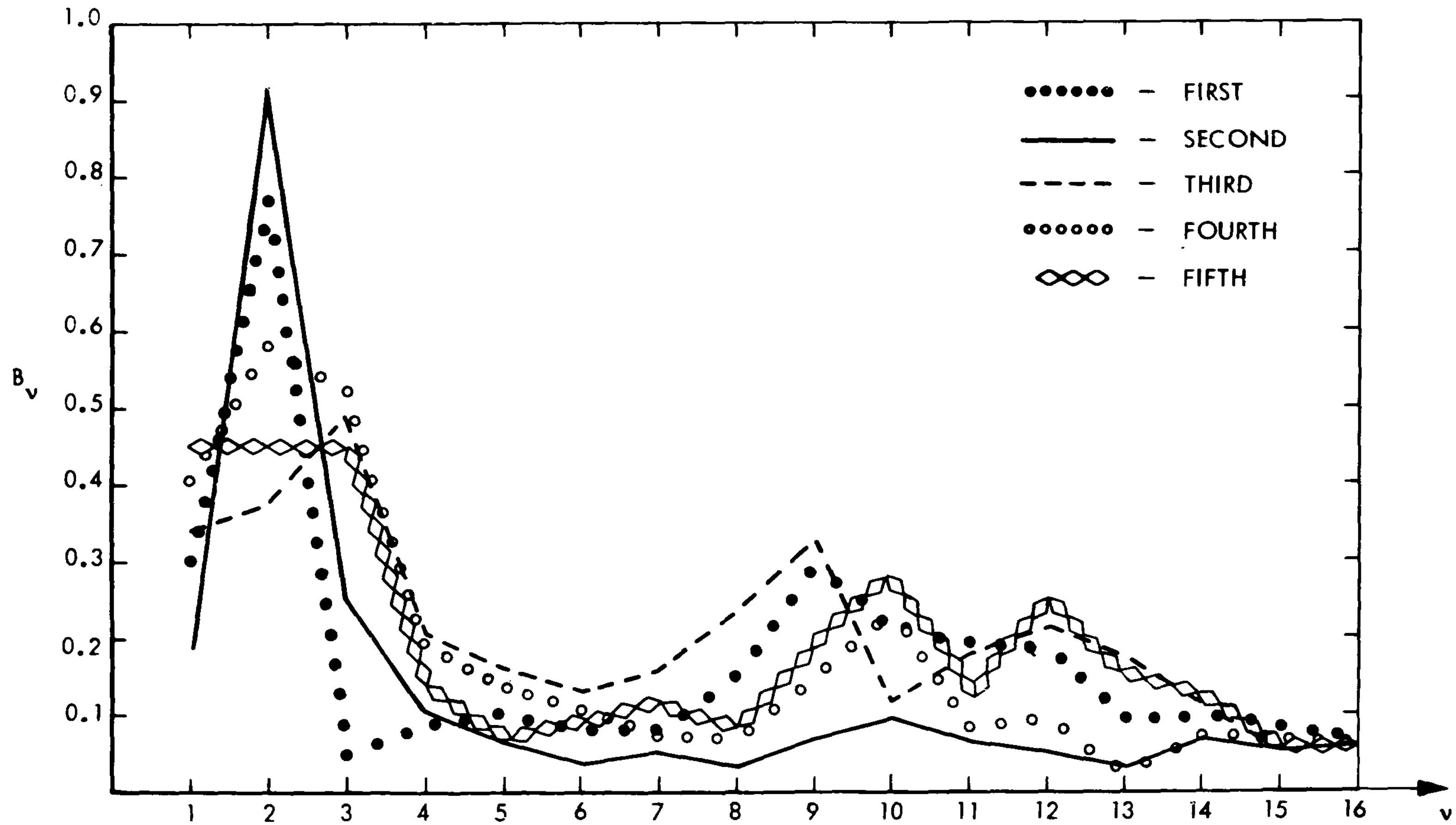


Figure 7. Orthonormal Spectral Pattern for hid (Vowel)

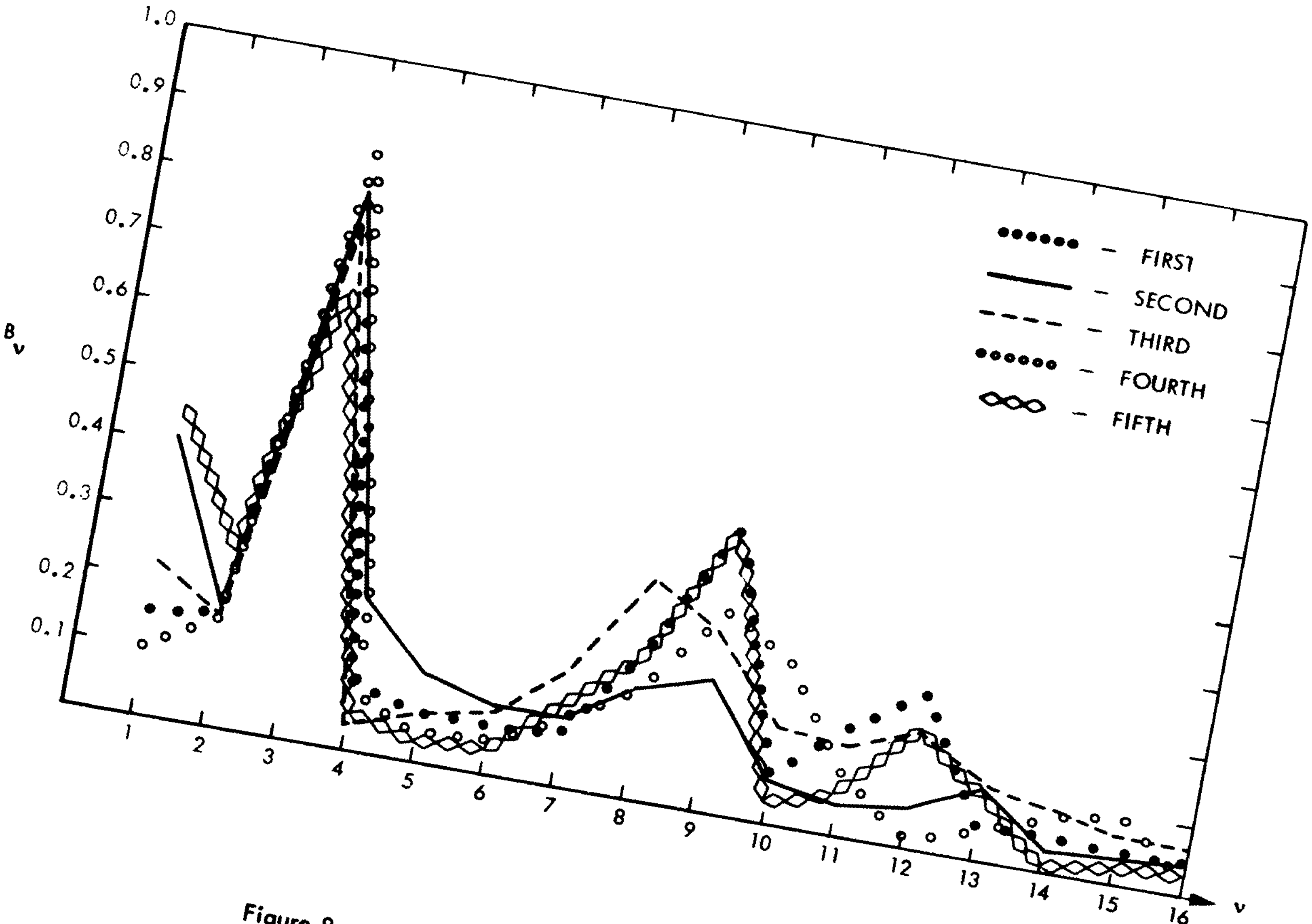


Figure 8. Orthonormal Spectral Pattern for <u>head</u> (Vowel)

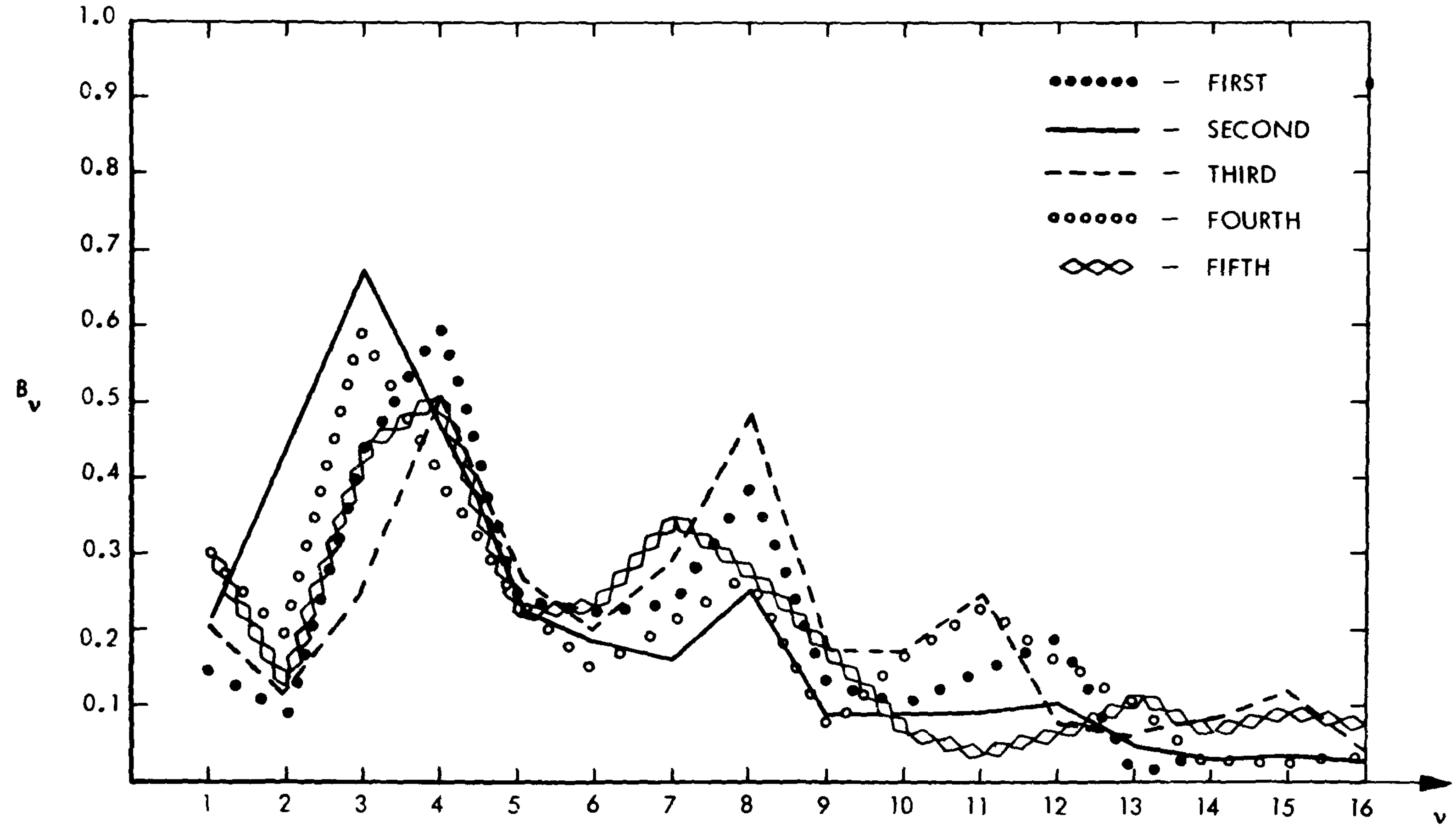


Figure 9. Orthonormal Spectral Pattern for h<u>a</u>d (Vowel)

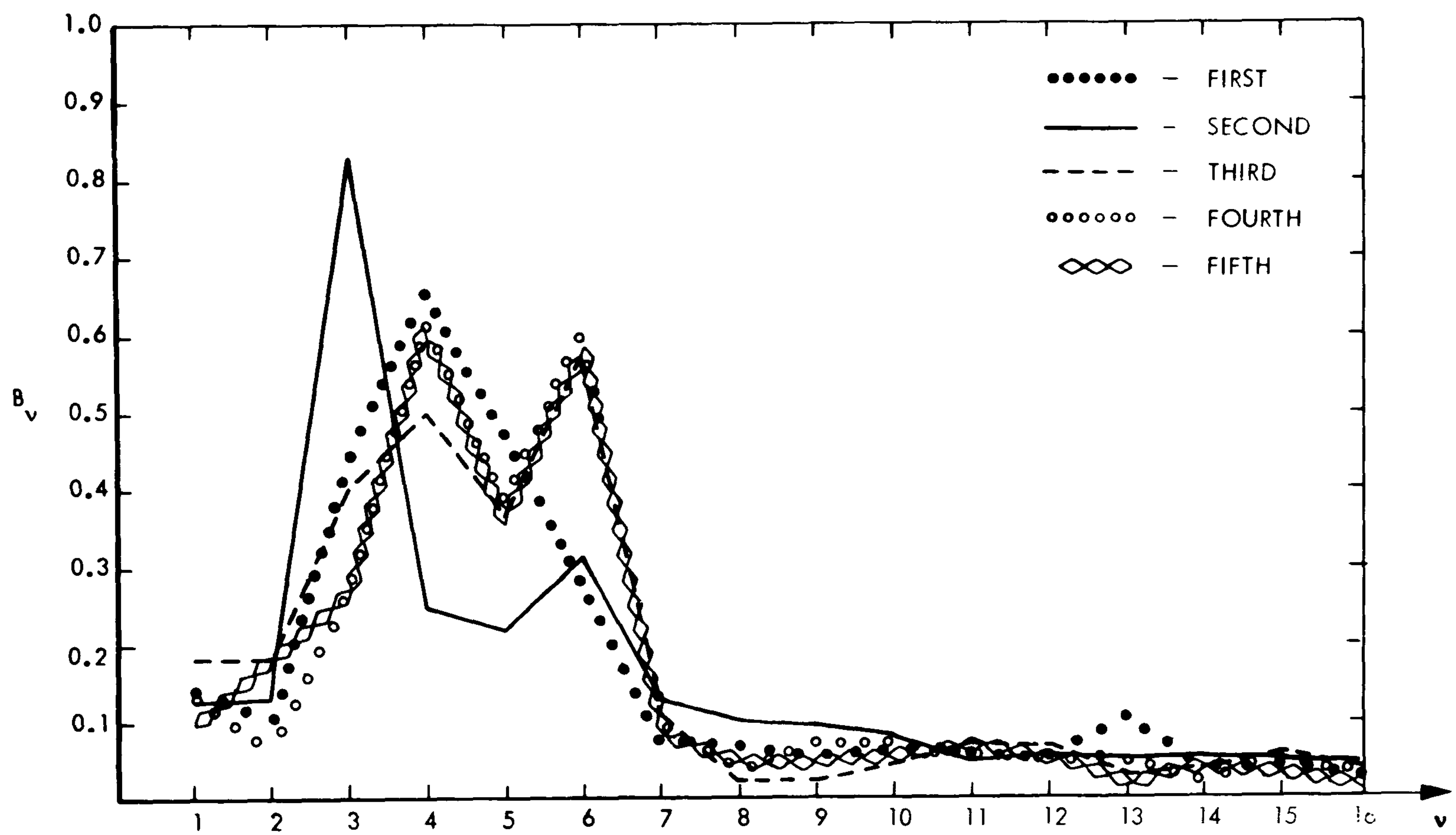


Figure 10. Orthonormal Spectral Pattern for h<u>u</u>b (Vowel)

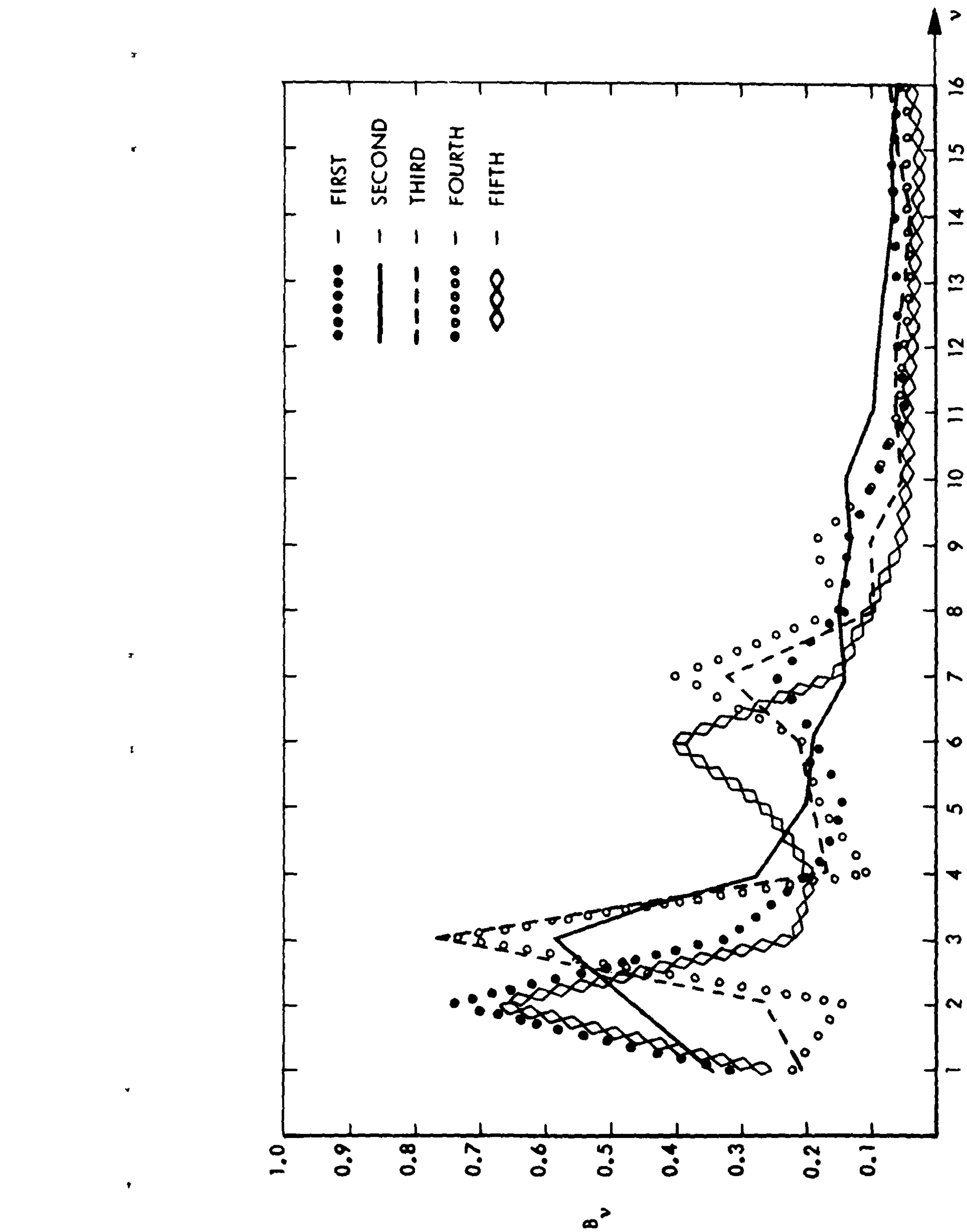


Figure 11. Orthonormal Spectral Pattern for her (Vowel)

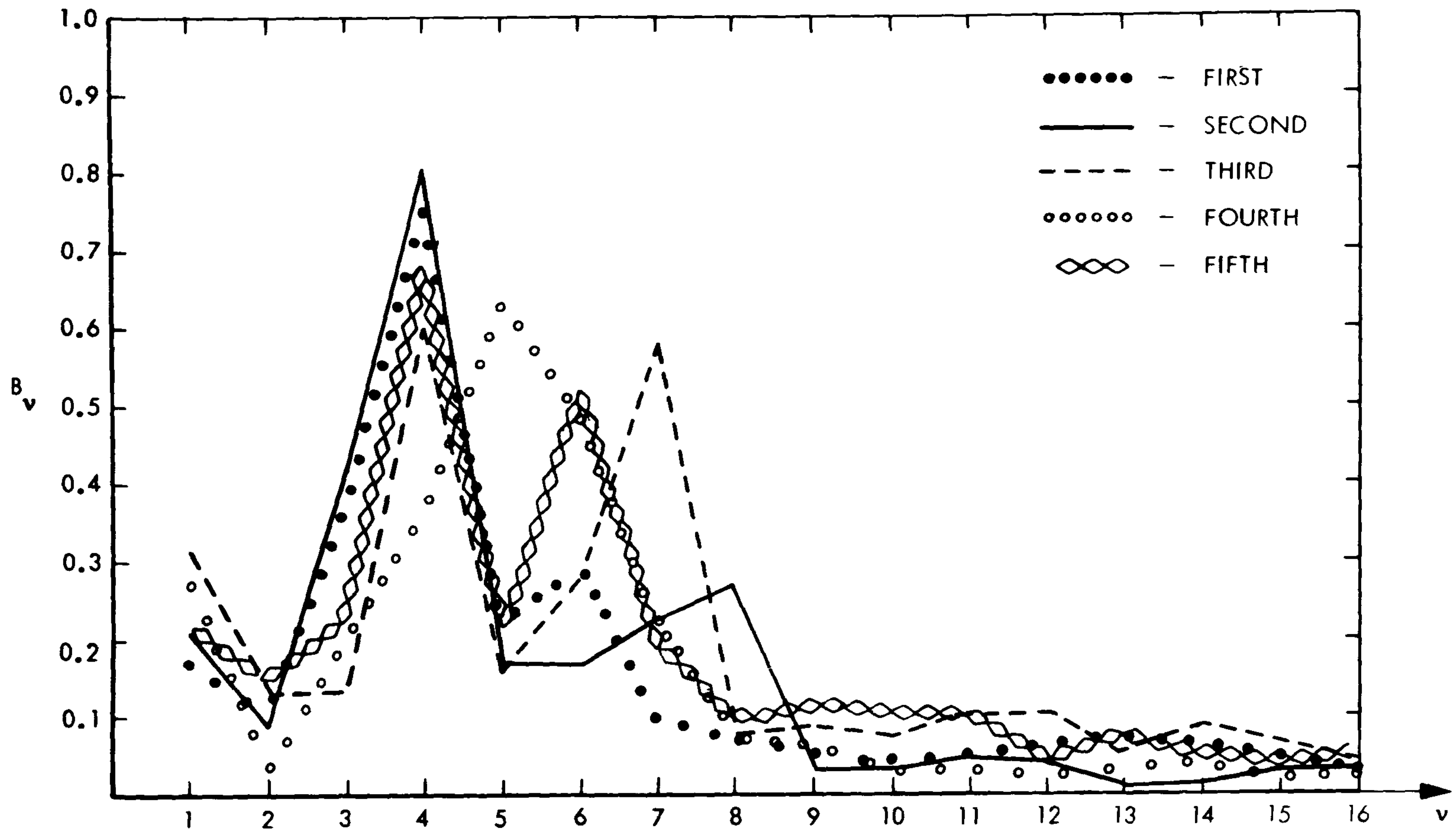


Figure 12. Orthonormal Spectral Pattern for ah (Vowel)

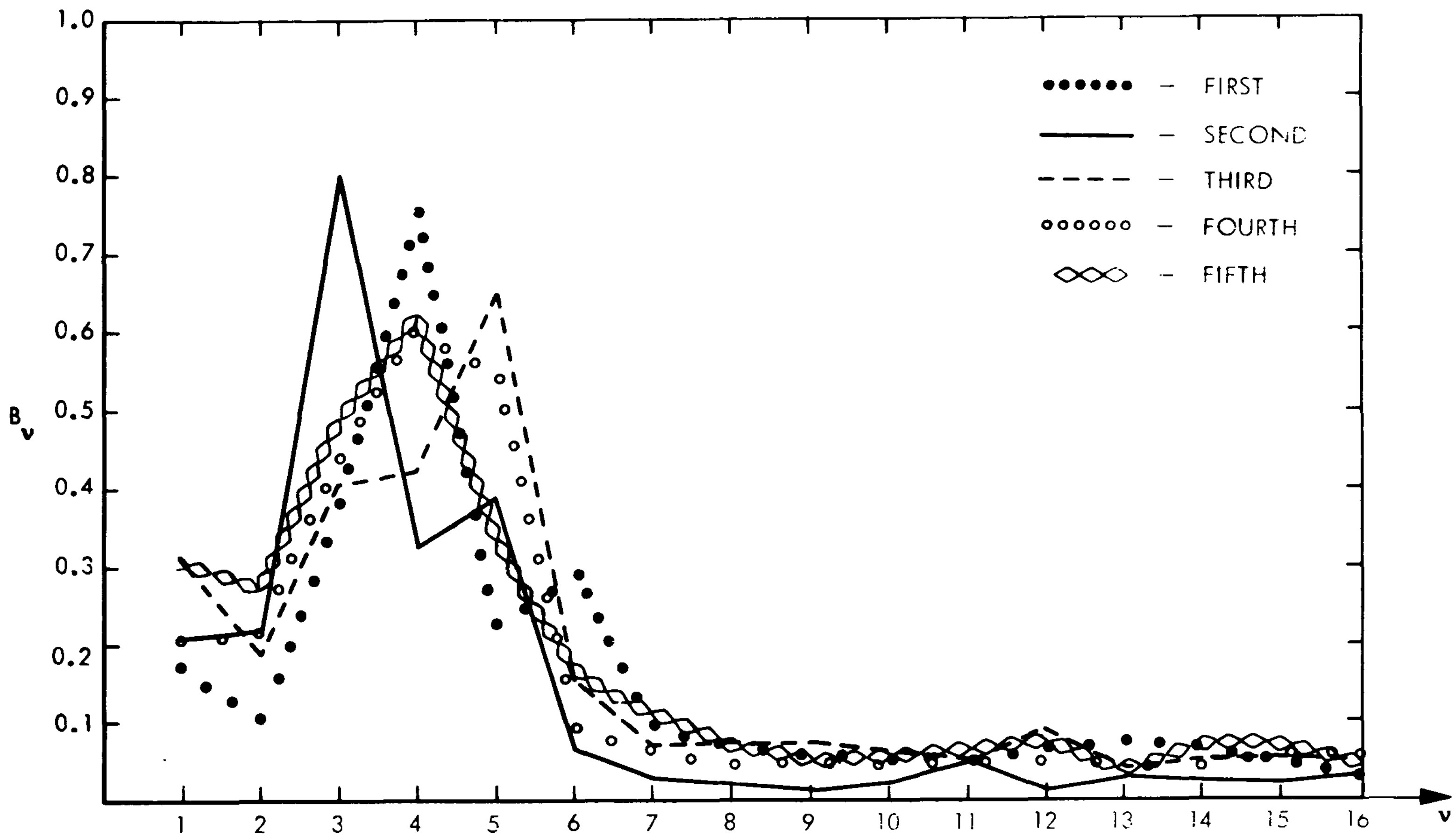


Figure 13. Orthonormal Spectral Pattern for <u>awe</u> (Vowel)

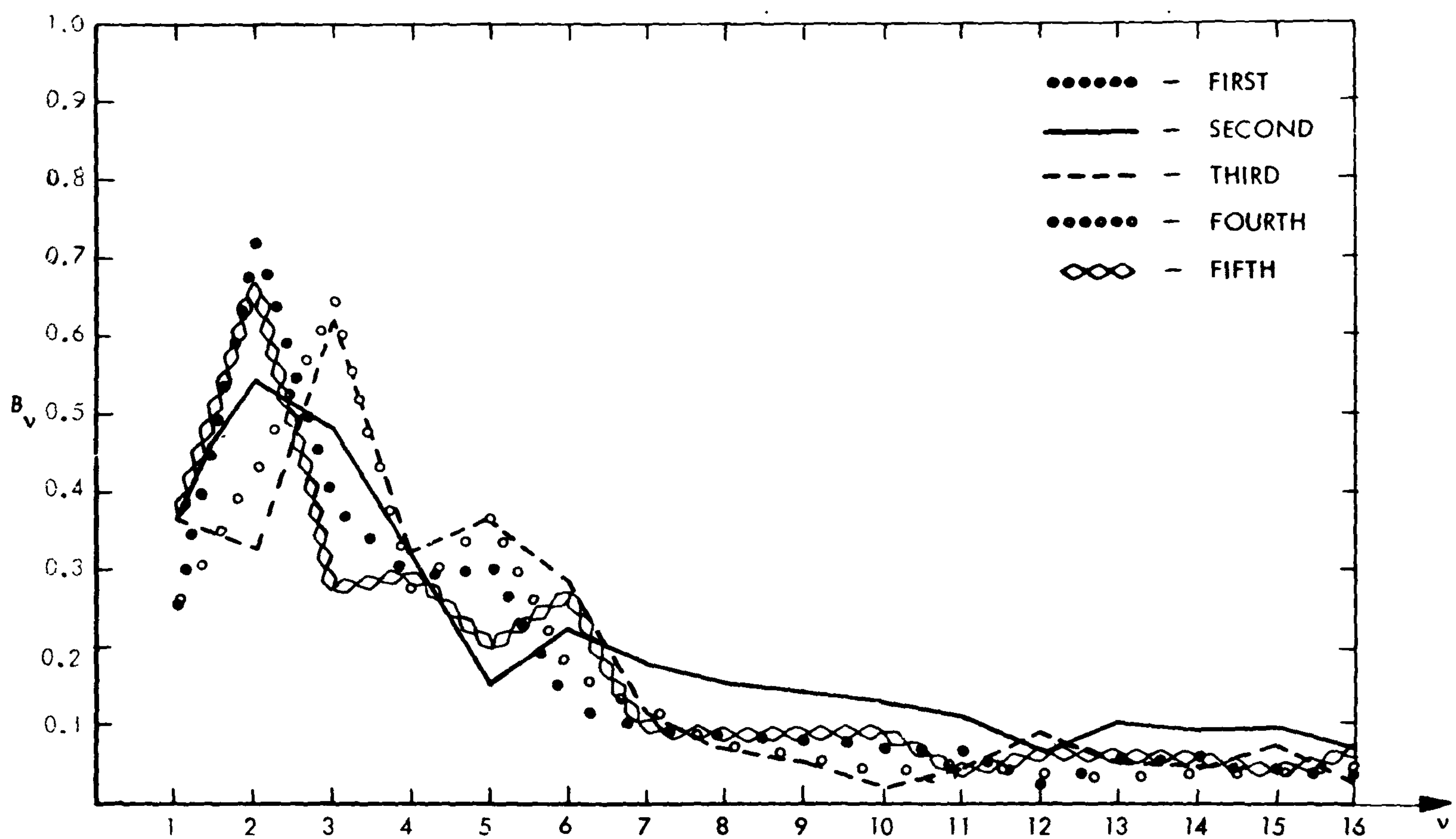


Figure 14. Orthonormal Spectral Pattern for h<u>oo</u>d (Vowel)

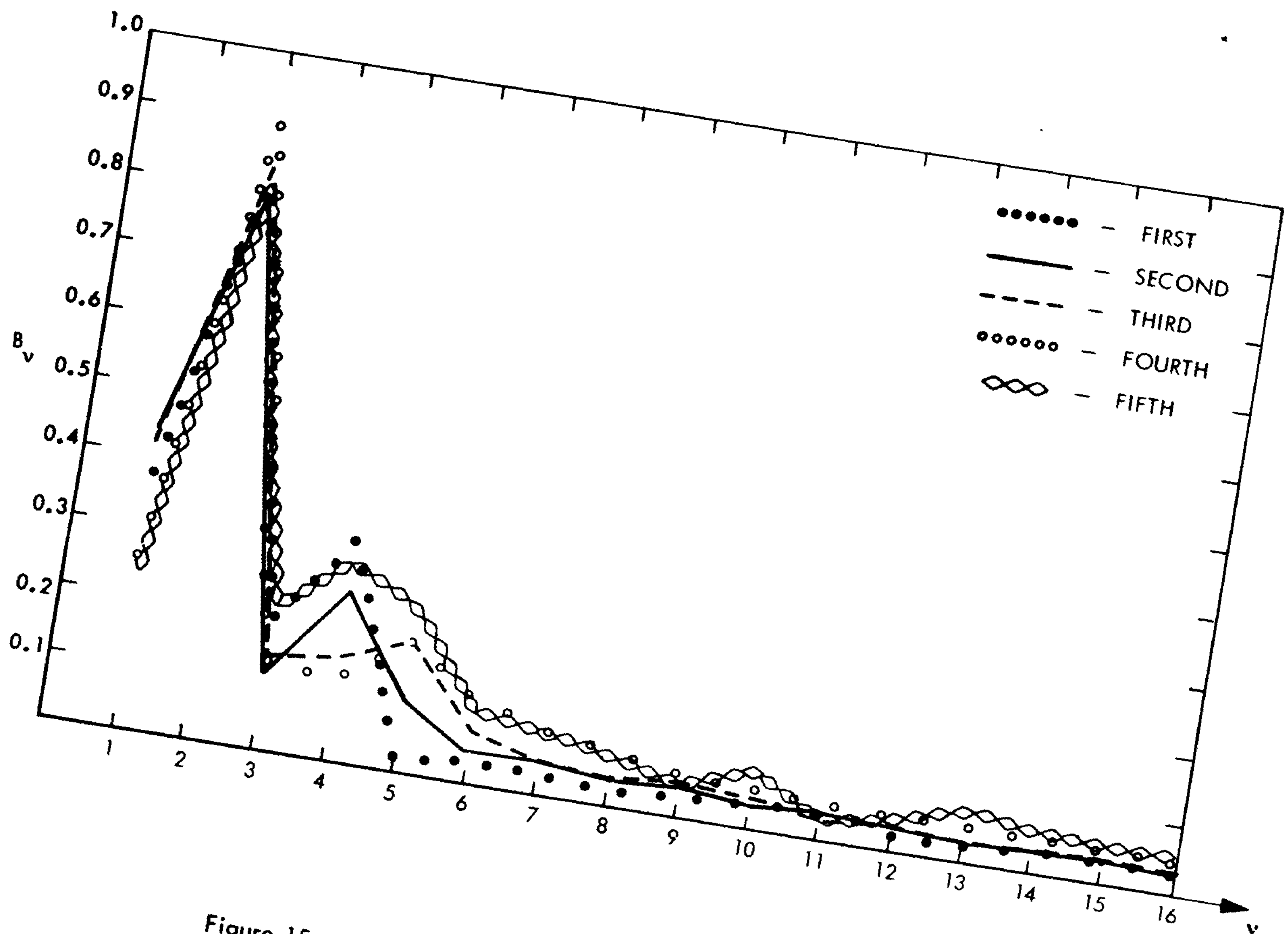


Figure 15. Orthonormal Spectral Pattern for who (Vowel)

The peaks in the patterns of Figures 6 through 15 are generally correlated with formant locations which are the basic parameters associated with the articulatory configuration for each vowel.

Speaker to speaker variations are, of course, also evidenced in the patterns shown. These variations, of course, always reduce the reliability of automatic speech recognition logics. On the other hand, the speaker related variations evidenced in Figures 6 through 15 may also be exploited for speaker recognition applications.

Of course, these data must be taken only as preliminary indications since an analysis of variance of large quantities of data (than could be collected within the scope of the current limited program) would be needed to produce definitive results as to the reliability of a recognition process using this type of data.

SECTION 3

CONCLUSIONS

The following conclusions can be drawn from the results of the current six month program of speech analysis research.

(1) The orthogonalized exponentially damped sinusoidal set using sixteen fixed frequencies and fixed damping at each frequency is a valid representation for pitch synchronous analysis of speech signals independent of the individual speaker.

(2) The above representation is efficient for speech waveform analysis in the sense that the experimentally derived cross-correlations between orthonormal coefficients are small.

(3) The exponentially damped sinusoidal representation is less efficient for speech energy spectra than for the waveform itself. This is evidenced by the fact that the cross-correlations among the phase eliminated coefficients are more highly cross-correlated than the basic orthonormal coefficients themselves. The spectrum representation can be made more efficient either by linear coding techniques or by modifying the analysis process so that the Fourier coefficients of the energy spectrum are derived directly by means of optimized autocorrelation analysis.

(4) The energy spectrum of the phase eliminated orthonormal coefficients derived by the method of Fourier analysis in terms of exponentially damped sinusoids has been found to be essentially limited to frequencies below 20 cycles/second. Thus, in the case of many speech sounds, it is feasible to time average the coefficient spectral data without destroying any of the fundamental phonemic information.

(5) The elimination of spectral phase information, in the generalized Fourier representation used in these studies, does not significantly reduce the intelligibility in an analysis-synthesis test. This process does, however, seriously reduce the quality of the resynthesized speech heard by a human listener.

(6) Theoretical studies have shown that it is possible to derive a new set of orthonormal functions, based on the orthogonalized exponentially damped sinusoids, wherein the coefficients will be linearly independent. Such a derived set of functions would consist of weighted sums of the original base functions and would be optimally efficient in the sense that the representation would provide a minimum expected mean square error for a given number of terms in the series.

(7) The "phase eliminated" spectrum patterns for ten vowels as uttered by five male speakers have been measured and these results indicate that it is feasible to perform either vowel recognition or speaker recognition using the orthonormal coefficients as input data. In the case of speaker recognition it is to be noted that all five speakers were of New England backgrounds.

REFERENCES

1. "Optimum Speech Signal Mapping Techniques," Final Technical Report No. F428-1, RADC-TR-62-3, Contract No. AF 30(602)-2446. Sylvania Electronic Systems, Applied Research Laboratory, 10 January 1962.

APPENDIX I

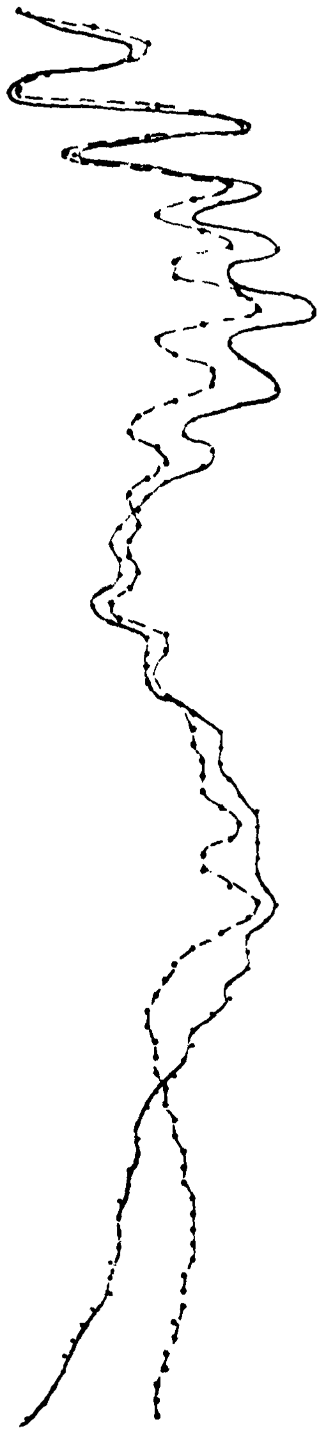
WAVEFORMS OF CENTRAL PITCH PERIODS IF TEN VOWEL UTTERANCES BY FIVE TALKERS RESYNTHESIZED IN TERMS OF FIXED ORTHOGONALIZED EXPONENTIALLY DAMPED SINUSOIDS

The fifty curves in Appendix I are designated in terms of file numbers, 1 through 5, and pitch period numbers, 1 through 10. The file number refers to the individual speaker and the pitch period number refers to a central pitch period of the vowel, portion of one of ten vowels. The five speakers are all mature males of New England state background. The words are designated as follows:

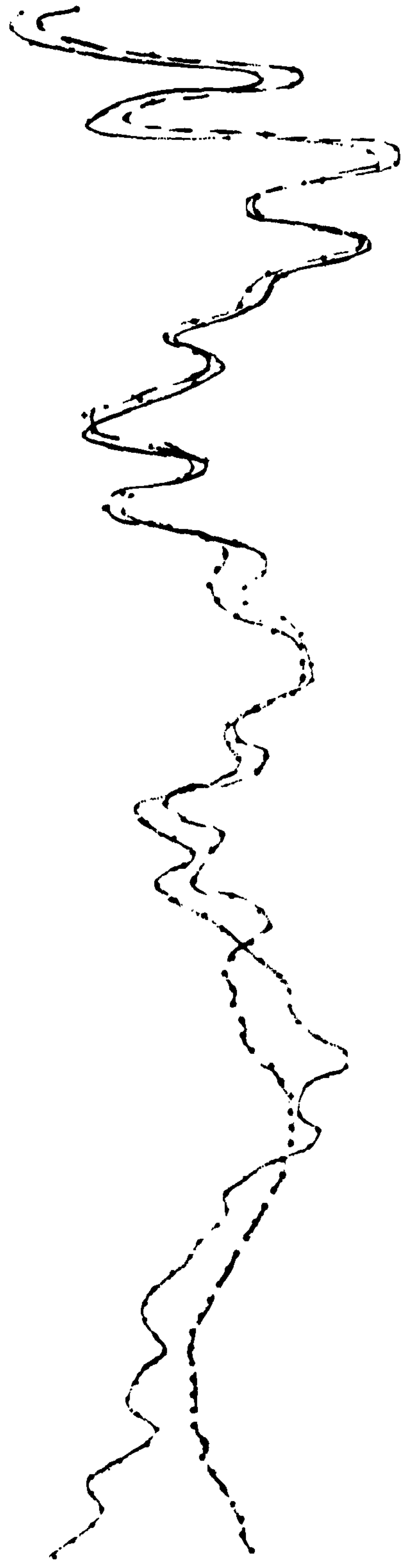
01	he
02	hid
03	head
04	had
05	hub
06	her
07	ah
08	awe
09	hood
10	who

Example: The Designation

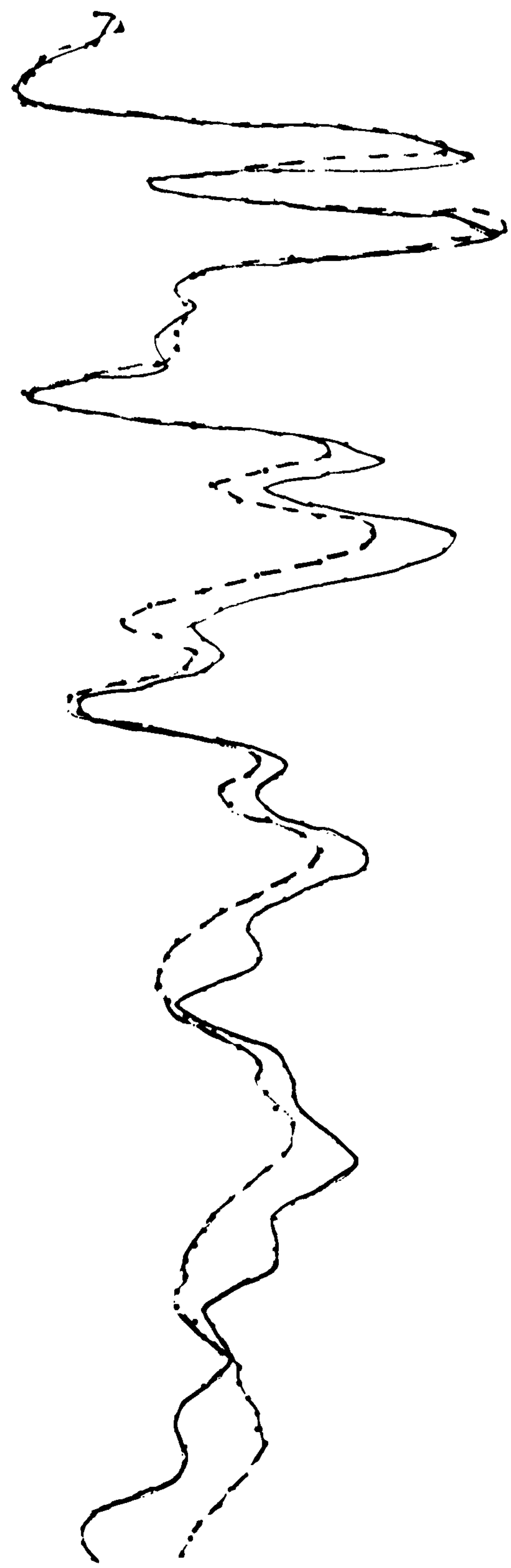
File number 05 pitch period number 01 refers to speaker number 5 and the vowel of the word he.



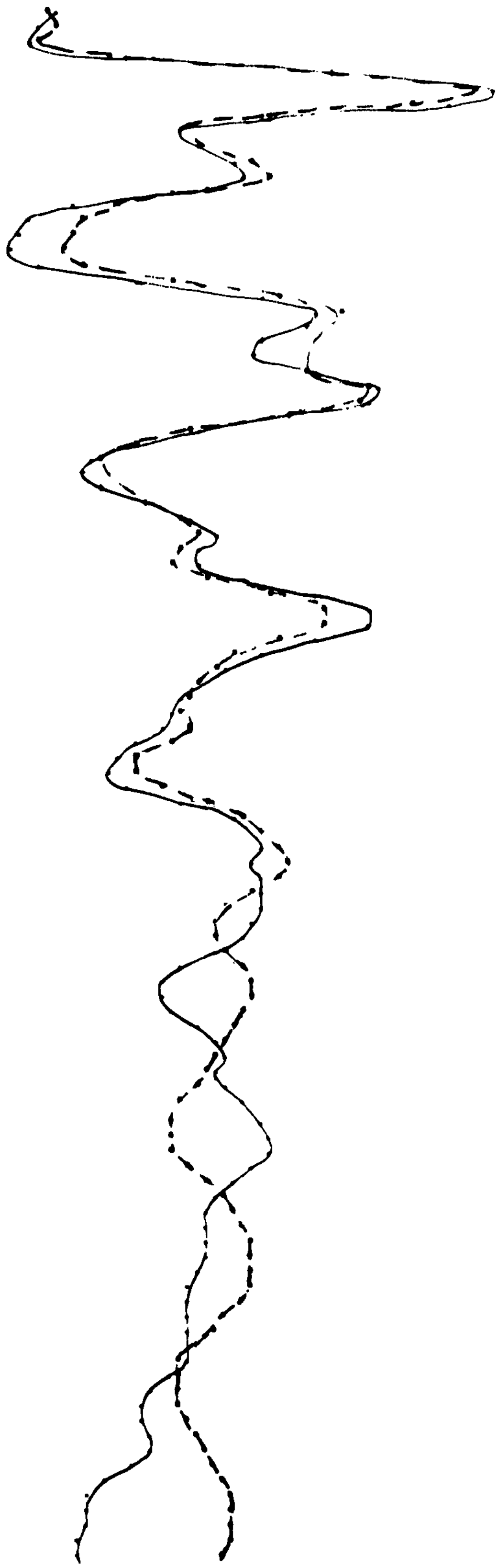
FILE NO. 1 PITCH PERIOD NO. 01 -- TIME OF FIRST PITCH POINT 010 SECONDS, 234 MILLISECONDS



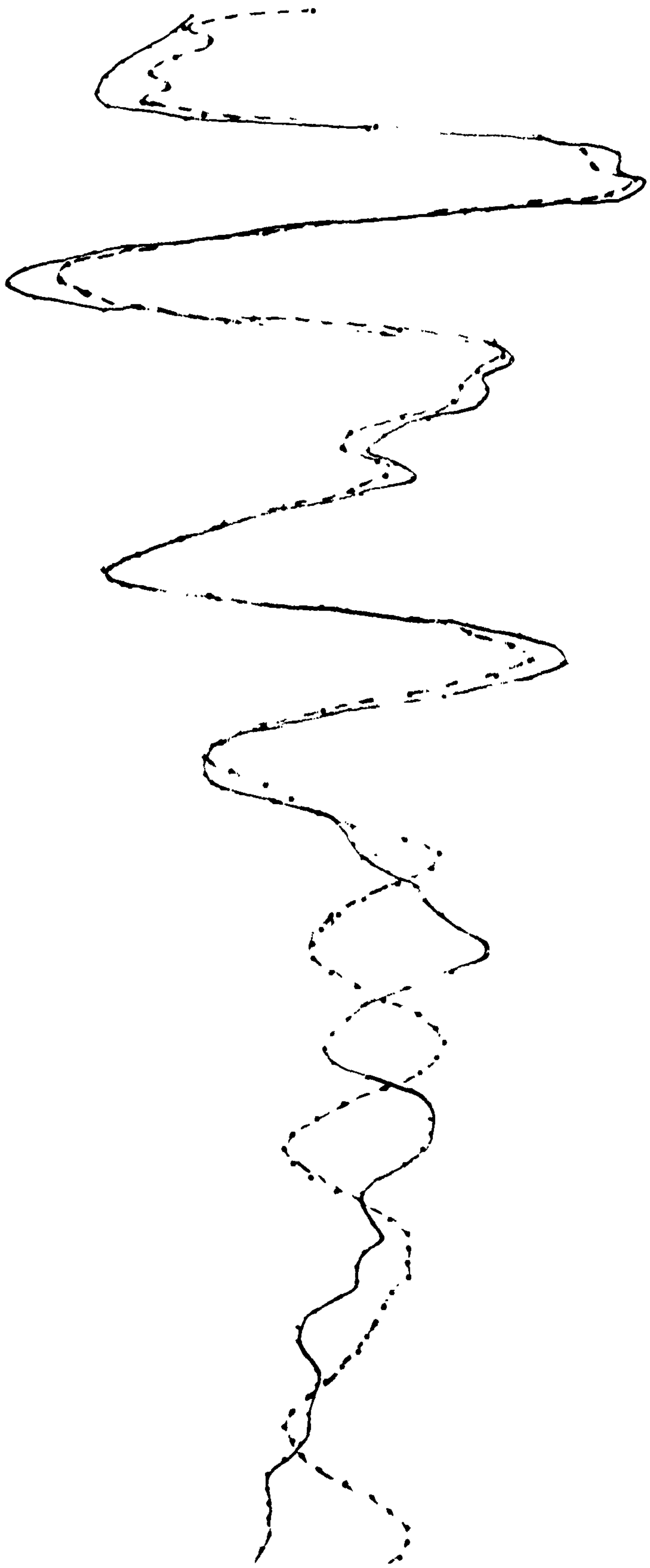
FILE NO. 1 PITCH PERIOD NO. 02 --TIME OF FIRST PITCH POINT 012 SECONDS, 779 MILLISECONDS



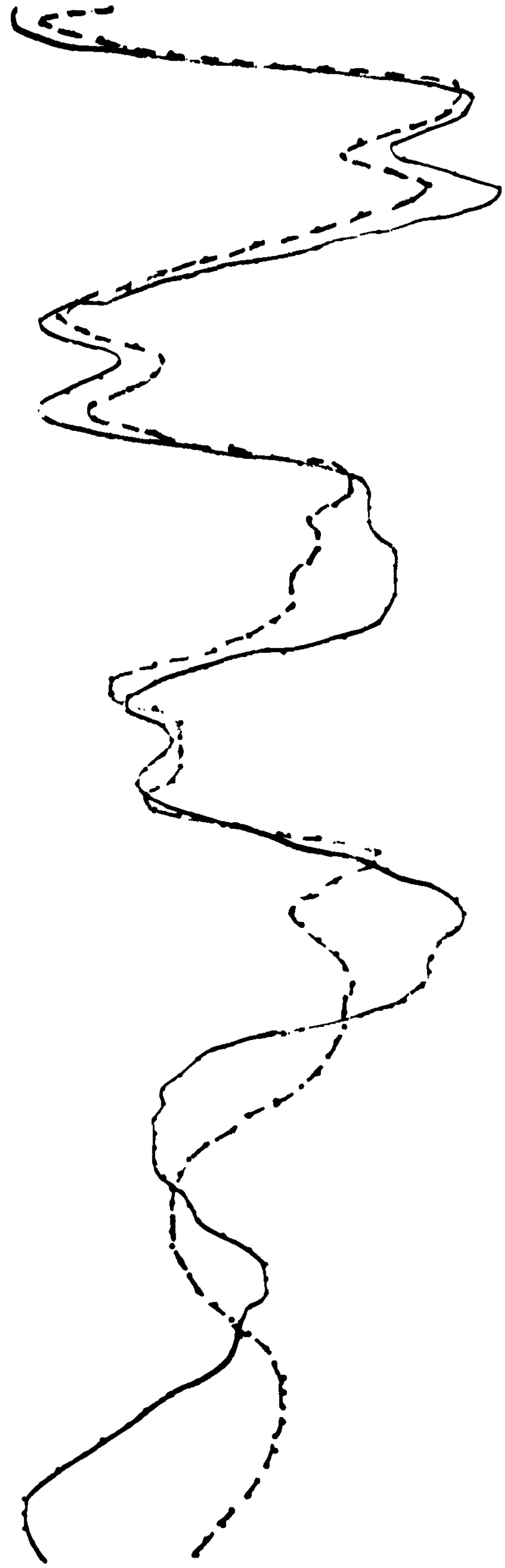
FILE NO. 1 PITCH PERIOD NO. 03 --TIME OF FIRST PITCH POINT 015 SECONDS, 524 MILLISECONDS



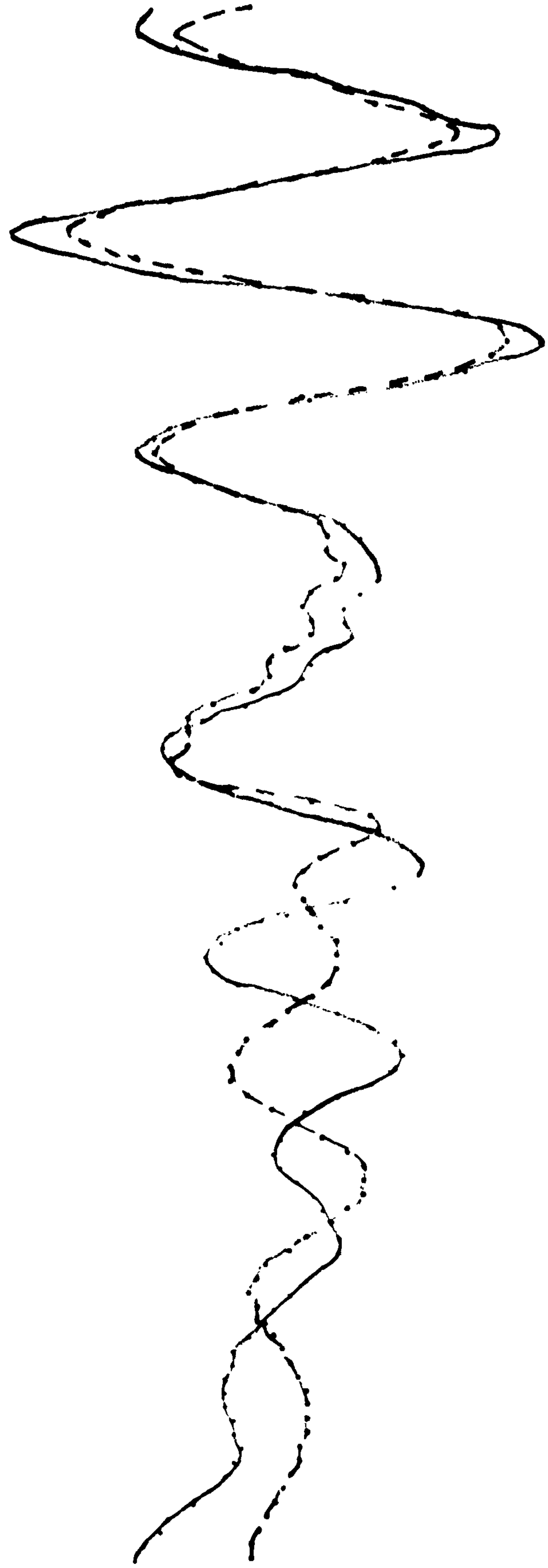
FILE NO. 1 PITCH PERIOD NO. 04 -- TIME OF FIRST PITCH POINT 018 SECONDS, 489 MILLISECONDS



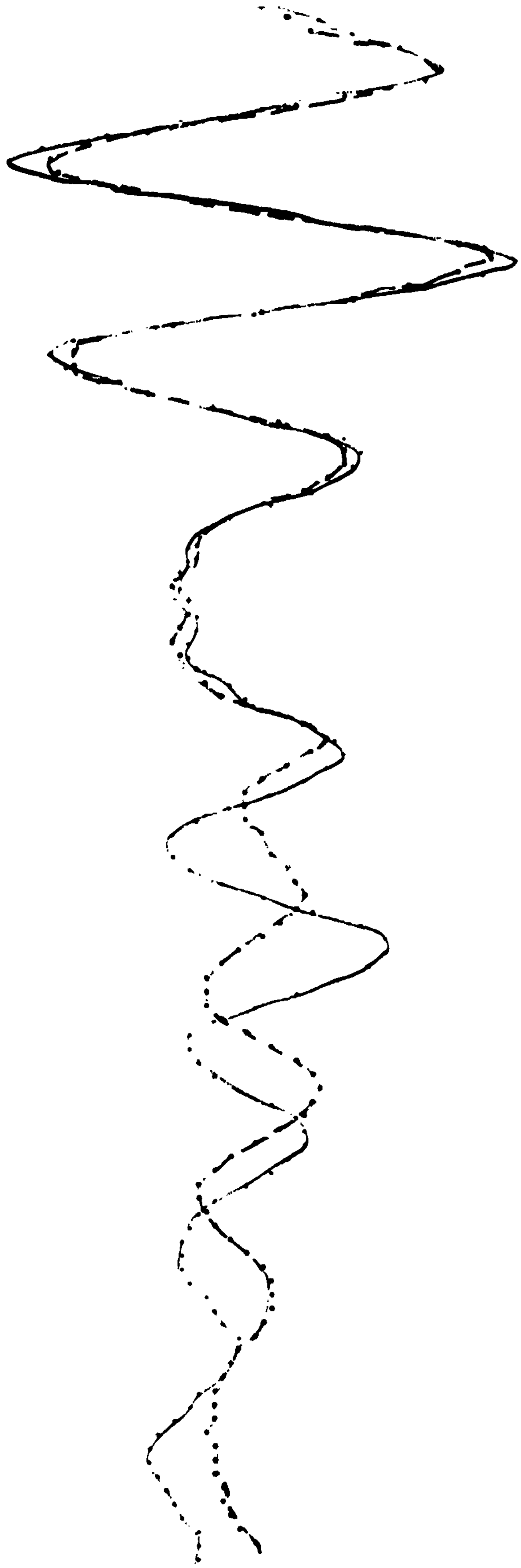
FILE NO. 1 PITCH PERIOD NO. 05 -- T' IRST PITCH POINT 021 SECONDS, 335 MILLISECONDS



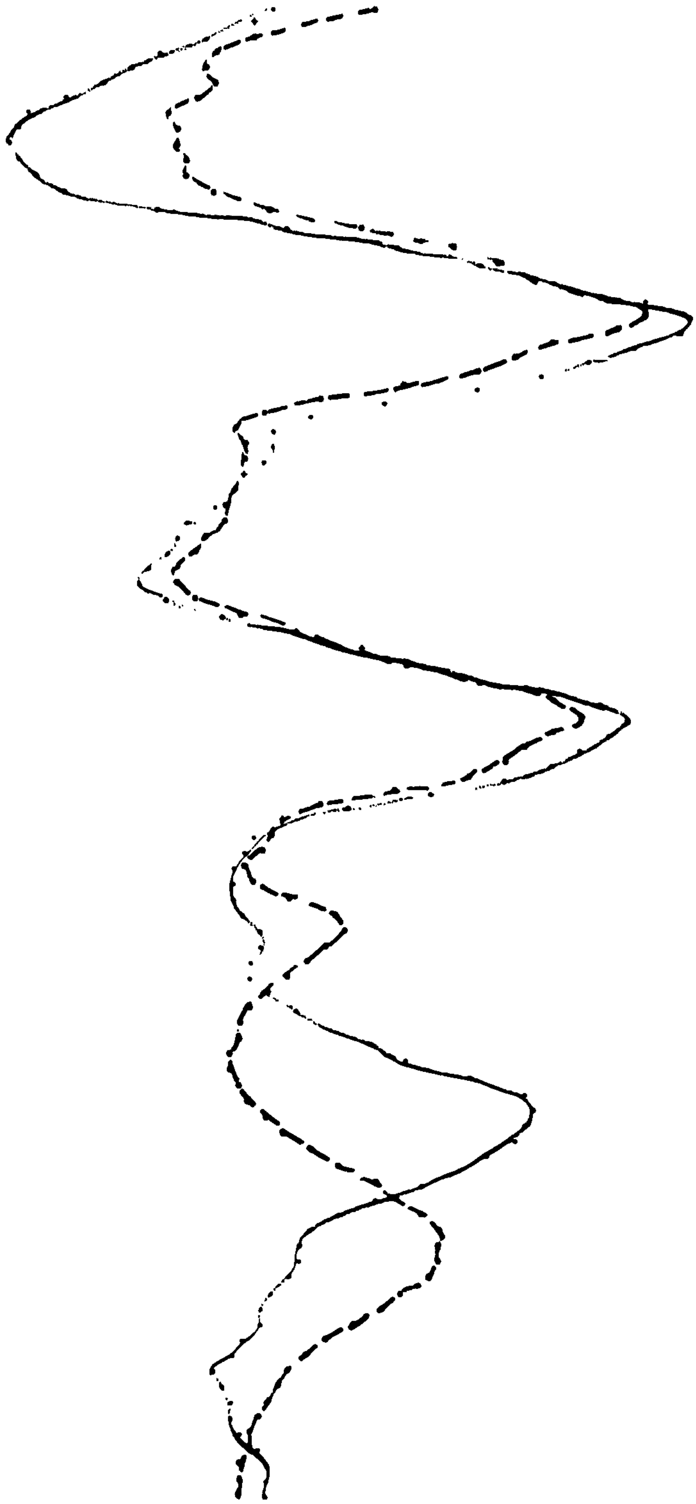
FILE NO. 1 PITCH PERIOD NO. 06 -- TIME OF FIRST PITCH POINT 024 SECONDS, 239 MILLISECONDS



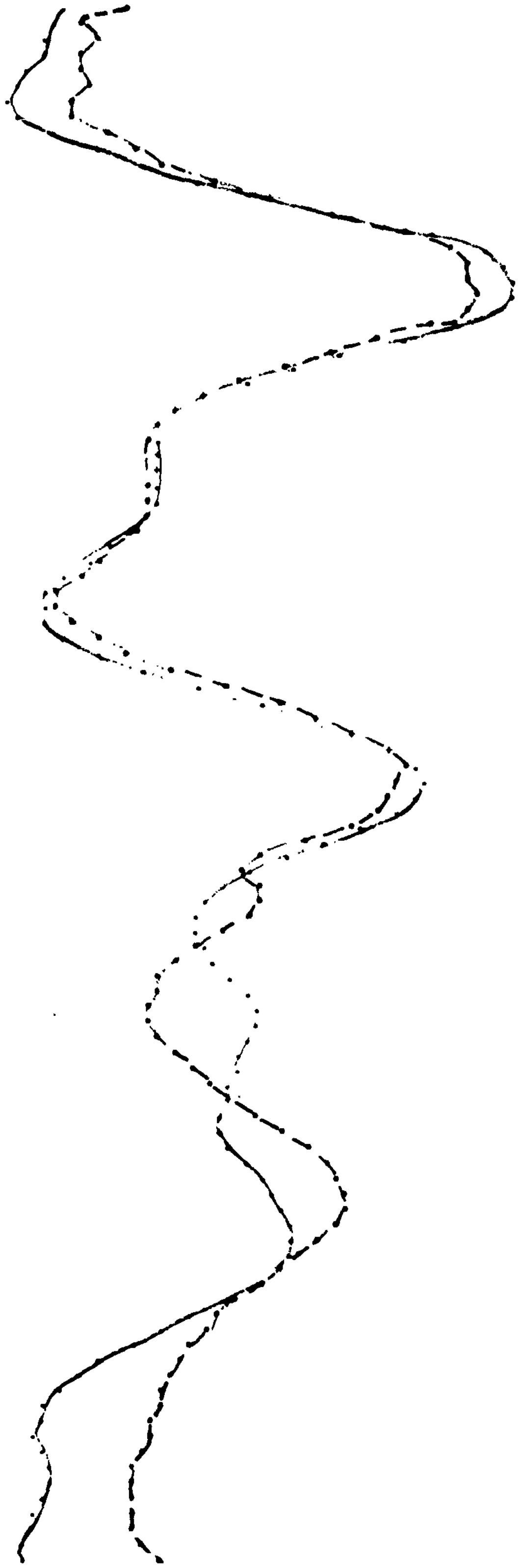
FILE NO. 1 PITCH PERIOD NO. 07 -- TIME OF FIRST PITCH POINT 027 SECONDS, 124 MILLISECONDS



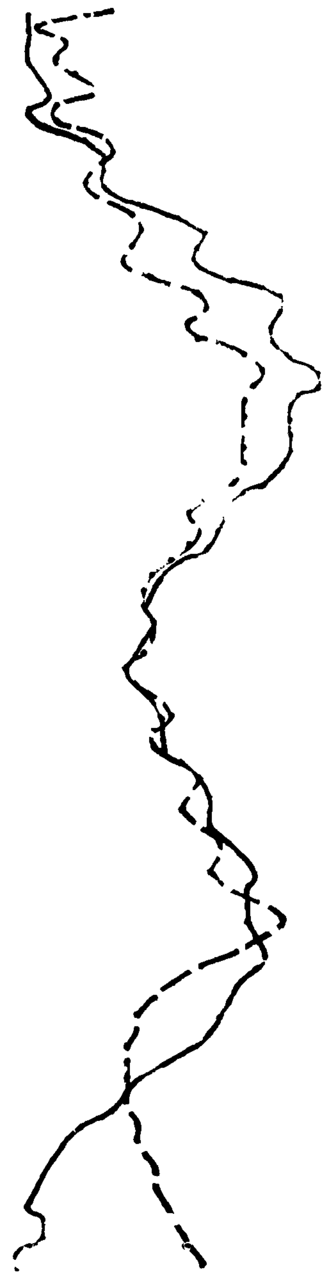
FILE NO. 1 PITCH PERIOD NO. 08 -- TIME OF FIRST PITCH POINT 030 SECONDS, 630 MILLISECONDS



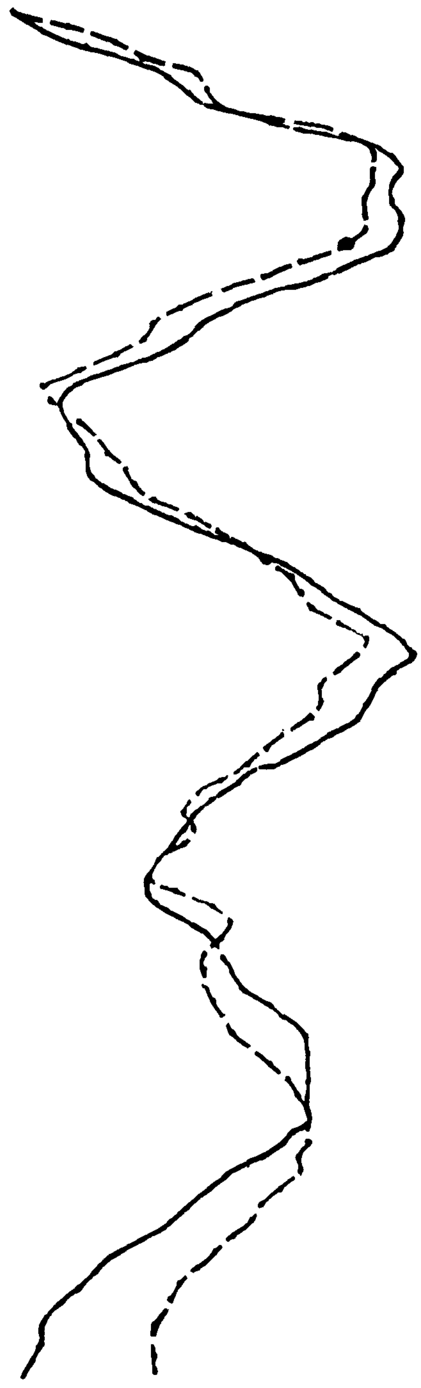
FILE NO. 1 PITCH PERIOD NO. 09 -- TIME OF FIRST PITCH POINT 033 SECONDS, 436 MILLISECONDS



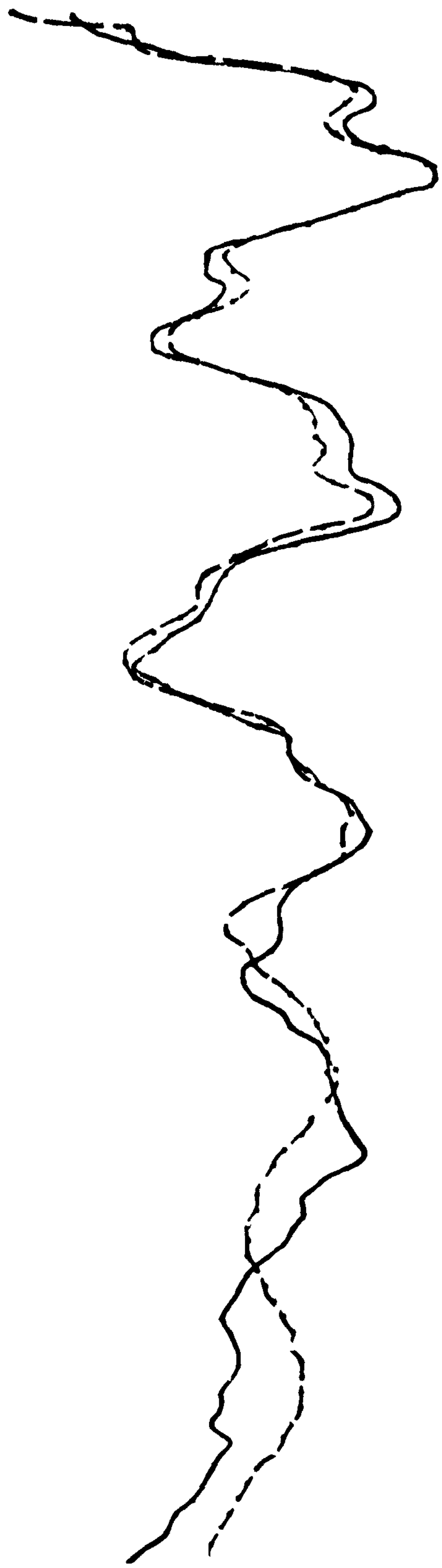
FILE NO. 1 PITCH PERIOD NO. 10 -- TIME OF FIRST PITCH POINT 036 SECONDS, 221 MILLISECONDS



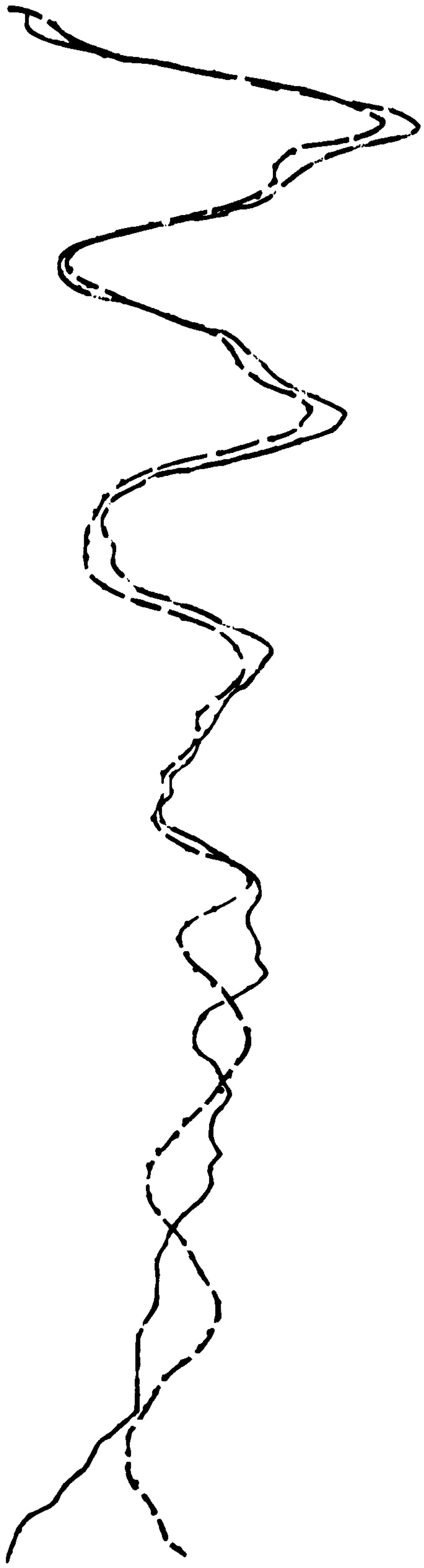
FILE NO. 2 PITCH PERIOD NO. 01 -- TIME OF FIRST PITCH POINT 028 SECONDS, 427 MILLISECONDS



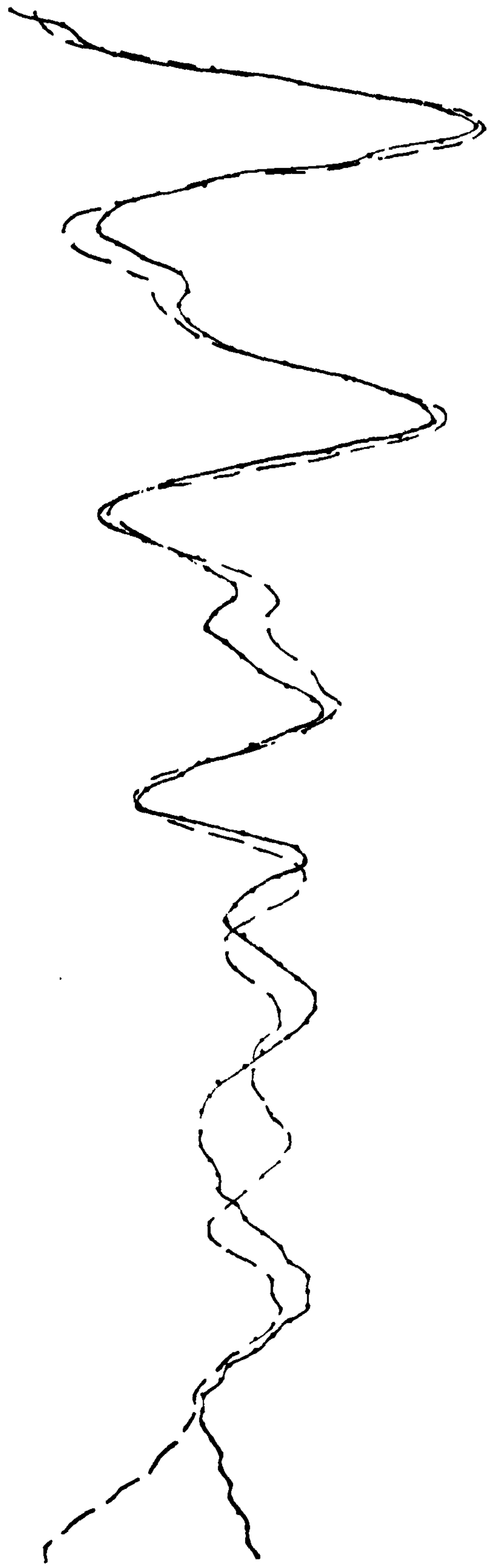
FILE NO. 2 PITCH PERIOD NO. 02 -- TIME OF FIRST PITCH POINT 030 SECONDS, 010 MILLISECONDS



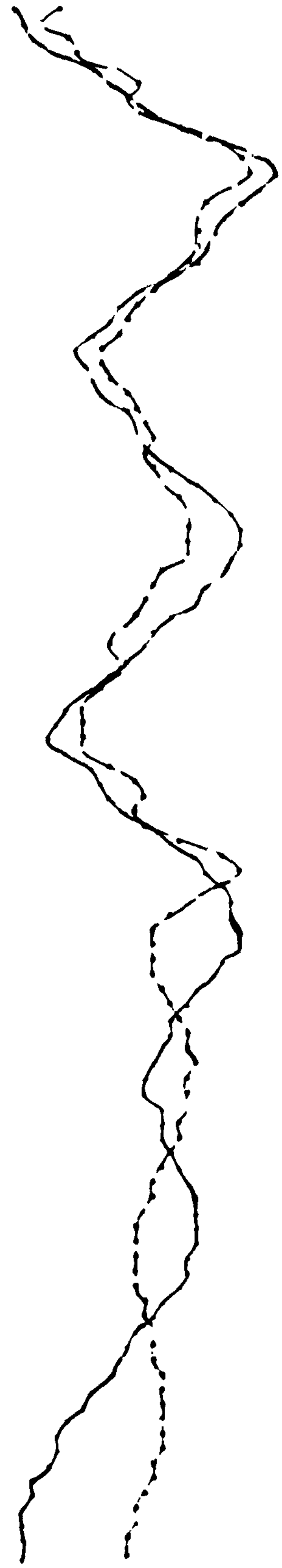
FILE NO. 2 PITCH PERIOD NO. 03 -- TIME OF FIRST PITCH POINT 031 SECONDS, 934 MILLISECONDS



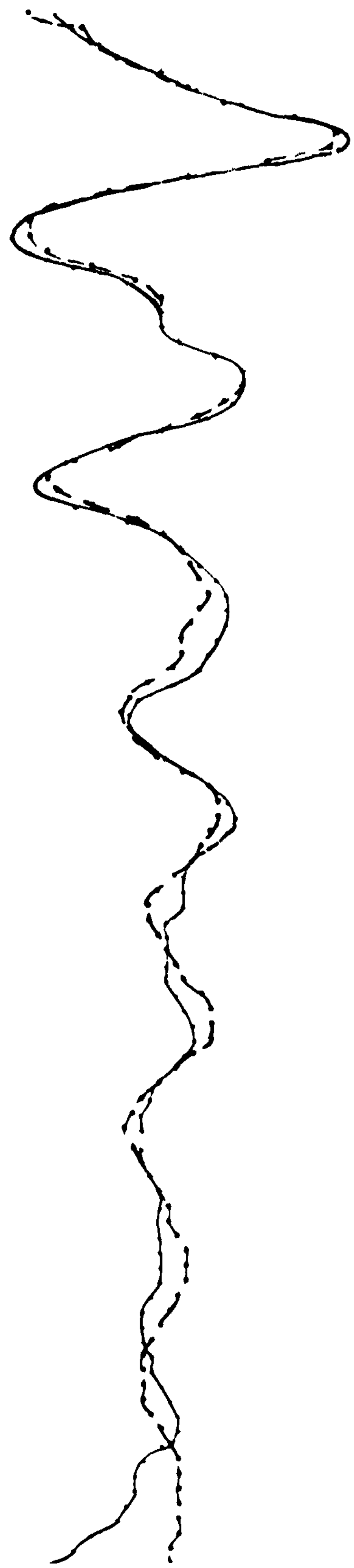
FILE NO. 2 PITCH PERIOD NO. 04 -- TIME OF FIRST PITCH POINT 033 SECONDS, 717 MILLISECONDS



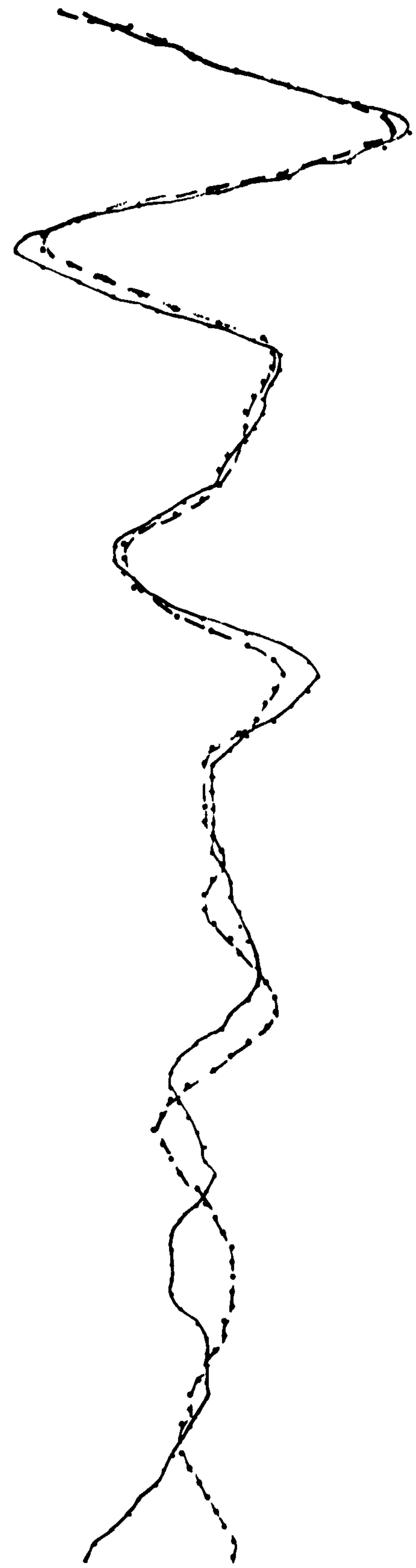
FILE NO. 2 PITCH PERIOD NO. 05 -- TIME OF FIRST PITCH POINT 035 SECONDS, 581 MILLISECONDS



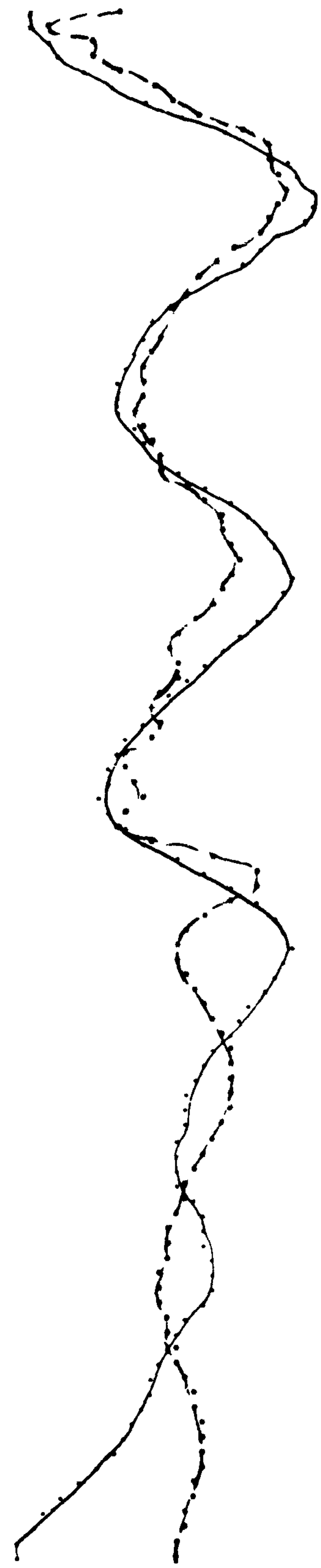
FILE NO. 2 PITCH PERIOD NO. 06 -- TIME OF FIRST PITCH POINT 037 SECONDS, 665 MILLISECONDS



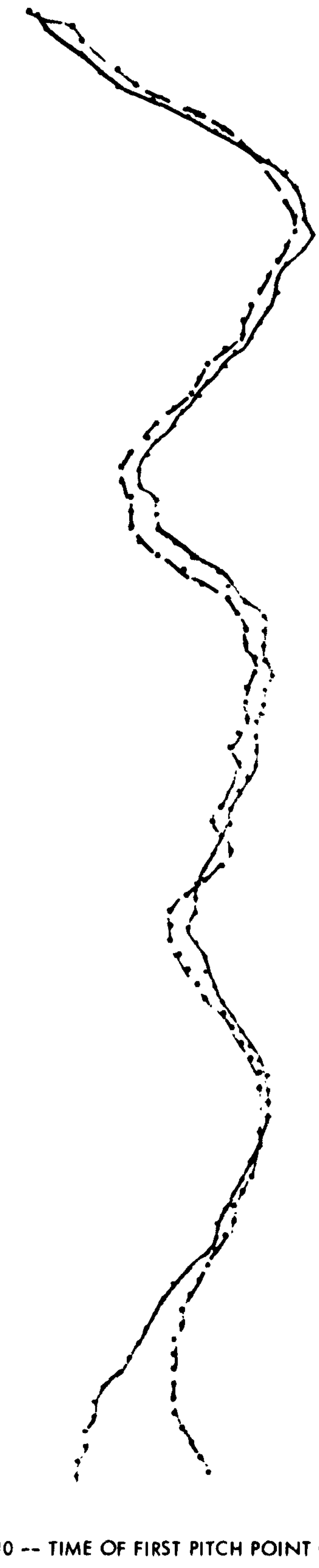
FILE NO. 2 PITCH PERIOD NO. 07 -- TIME OF FIRST PITCH POINT 039 SECONDS, 689 MILLISECONDS



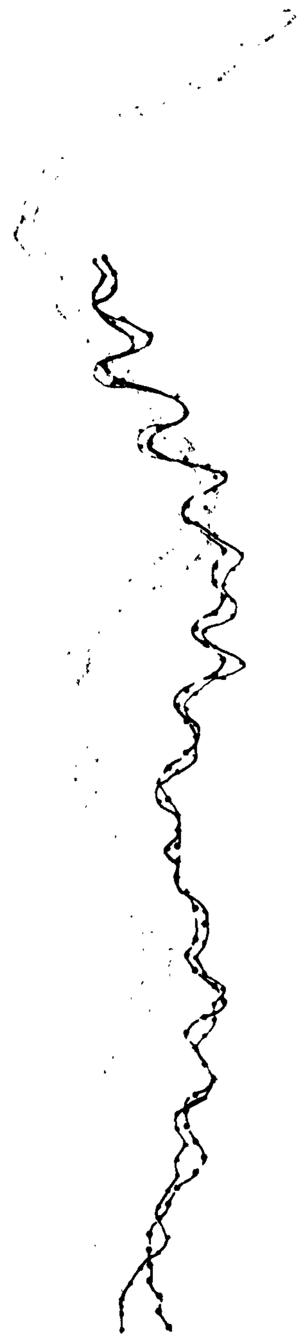
FILE NO. 2 PITCH PERIOD NO. 08 -- TIME OF FIRST PITCH POINT 041 SECONDS, 553 MILLISECONDS



FILE NO. 2 PITCH PERIOD NO. 09 -- TIME OF FIRST PITCH POINT 043 SECONDS, 537 MILLISECONDS

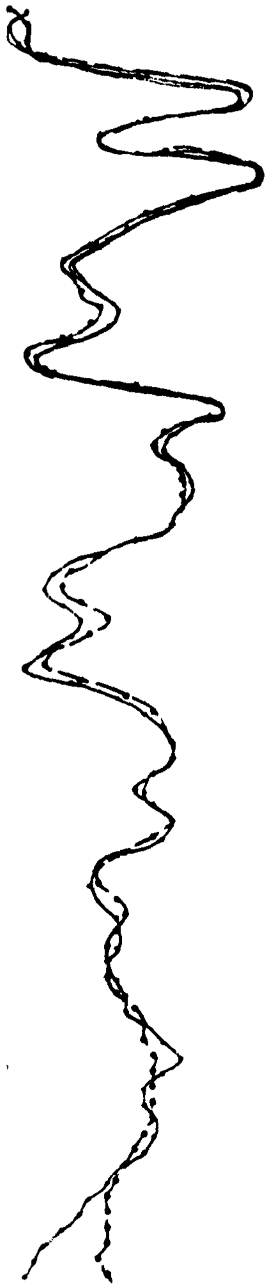


FILE NO. 2 PITCH PERIOD NO. 10 -- TIME OF FIRST PITCH POINT 045 SECONDS, 441 MILLISECONDS

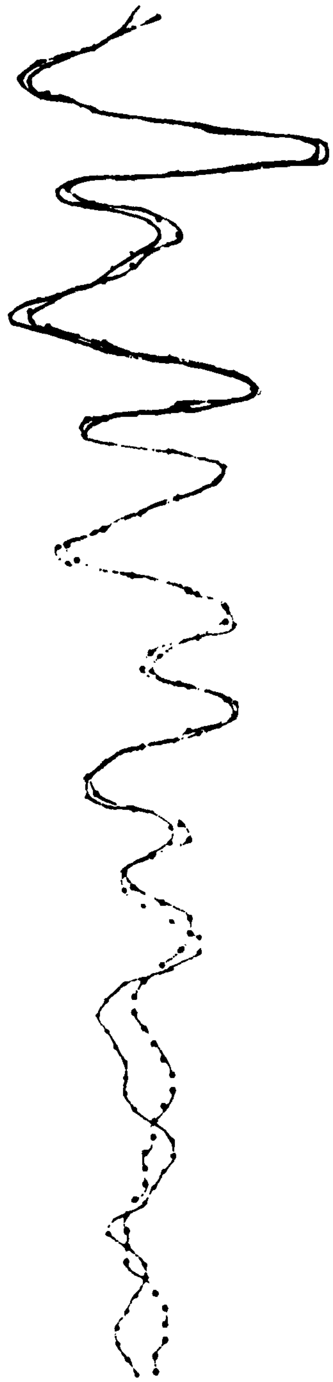


FILE NO. 3 PITCH PERIOD NO. 01 -- TIME OF FIRST PITCH POINT 020 SECONDS, 864 MILLISECONDS

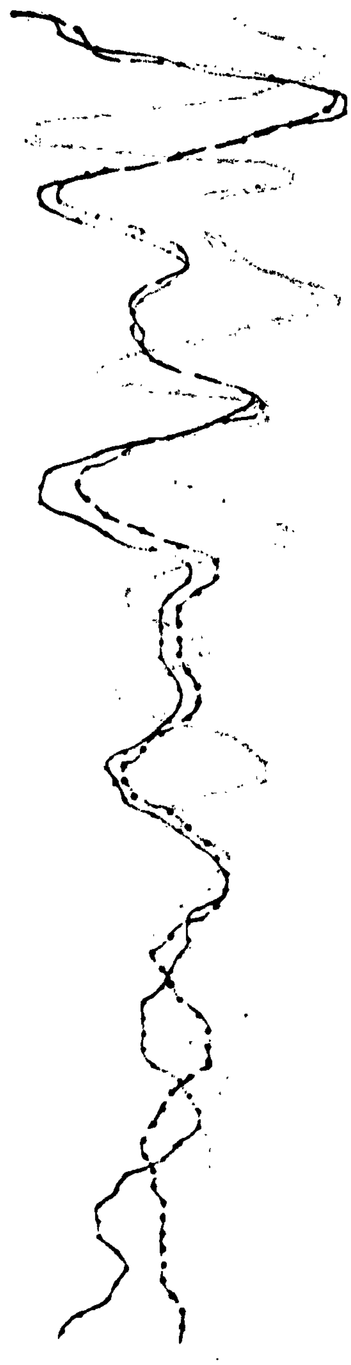
FILE NO. 3 PITCH PERIOD NO. 02 -- TIME OF FIRST PITCH POINT 022 SECONDS, 907 MILLISECONDS



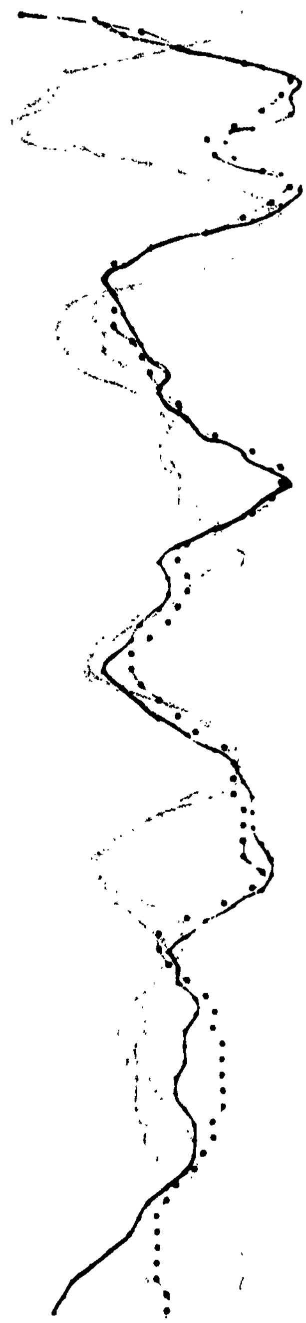
FILE NO. 3 PITCH PERIOD NO. 03 -- TIME OF FIRST PITCH POINT 025 SECONDS, 251 MILLISECONDS



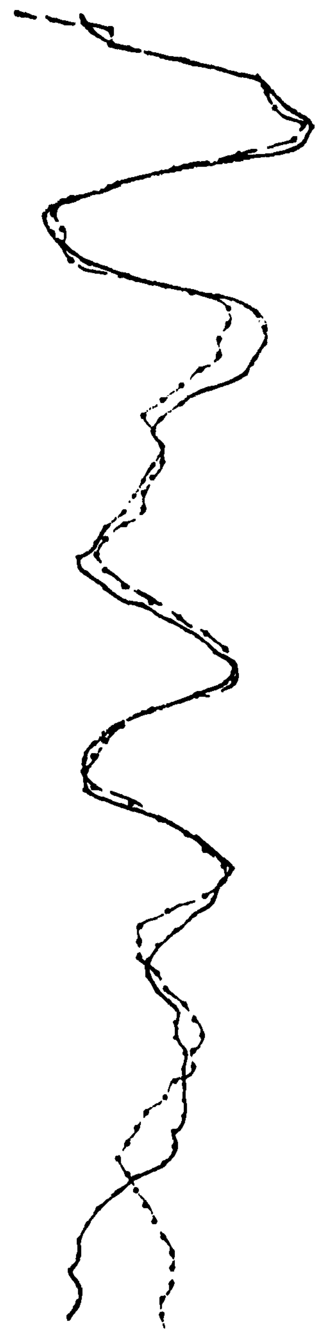
FILE NO. 3 PITCH PERIOD NO. 04 -- TIME OF FIRST PITCH POINT 027 SECONDS, 753 MILLISECONDS



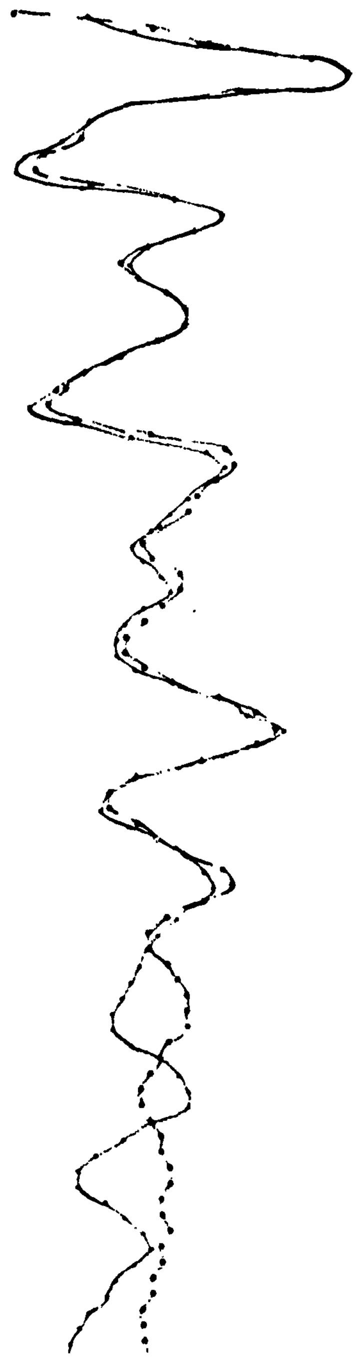
FILE NO. 3 PITCH PERIOD NO. 05 -- TIME OF FIRST PITCH POINT 030 SECONDS, 096 MILLISECONDS



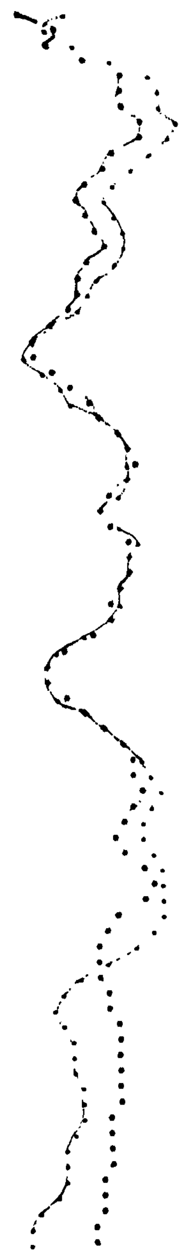
FILE NO. 3 PITCH PERIOD NO. 06 -- TIME OF FIRST PITCH POINT 032 SECONDS, 900 MILLISECONDS



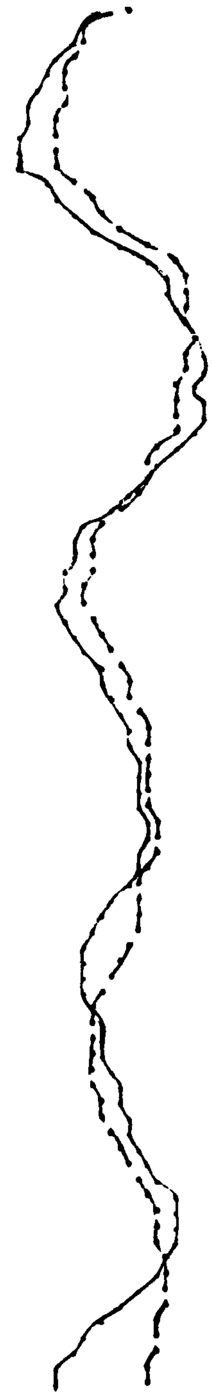
FILE NO. 3 PITCH PERIOD NO. 07 -- TIME OF FIRST PITCH POINT 036 SECONDS, 647 MILLISECONDS



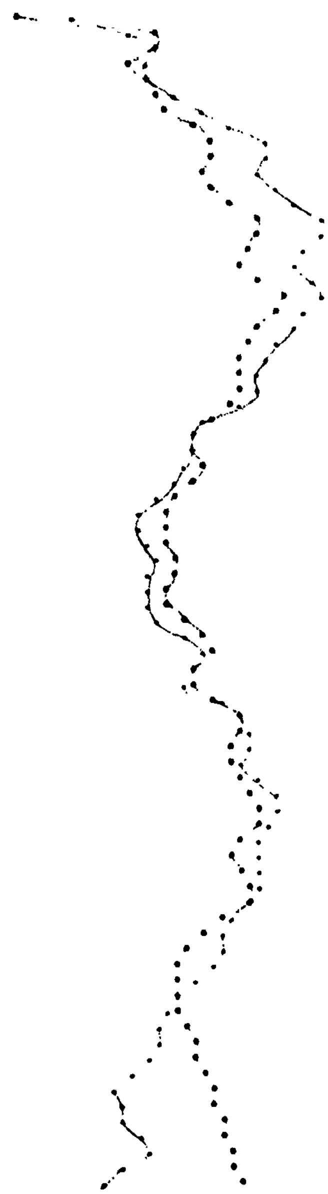
FILE NO. 3 PITCH PERIOD NO. 08 -- TIME OF FIRST PITCH POINT 041 SECONDS, 937 MILLISECONDS



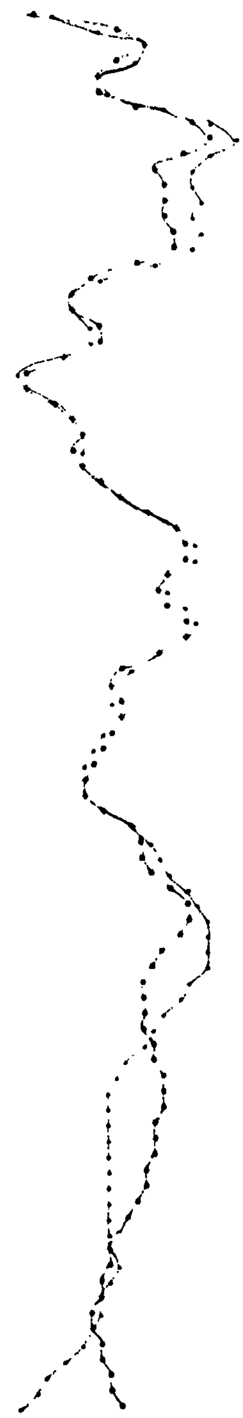
FILE NO. 3 PITCH PERIOD NO. 09 -- TIME OF FIRST PITCH POINT 044 SECONDS, 597 MILLISECONDS



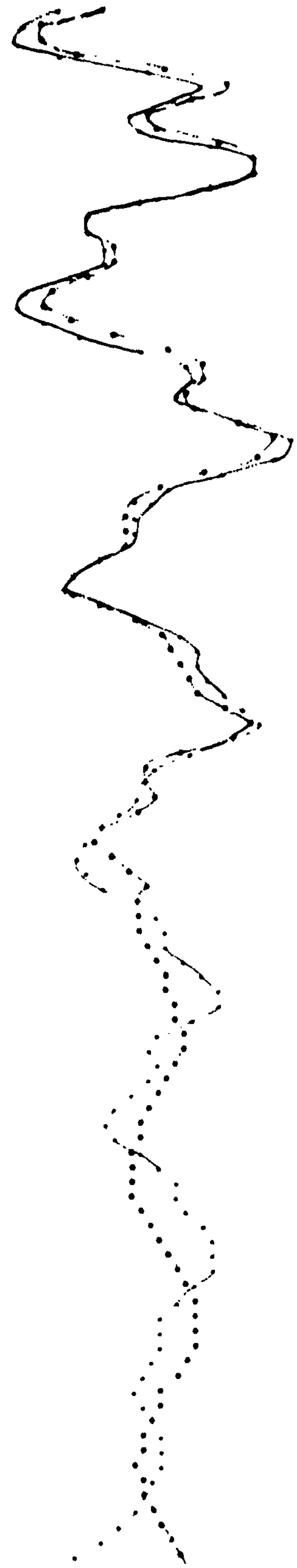
FILE NO. 3 PITCH PERIOD NO. 10 -- TIME OF FIRST PITCH POINT 046 SECONDS, 940 MILLISECONDS



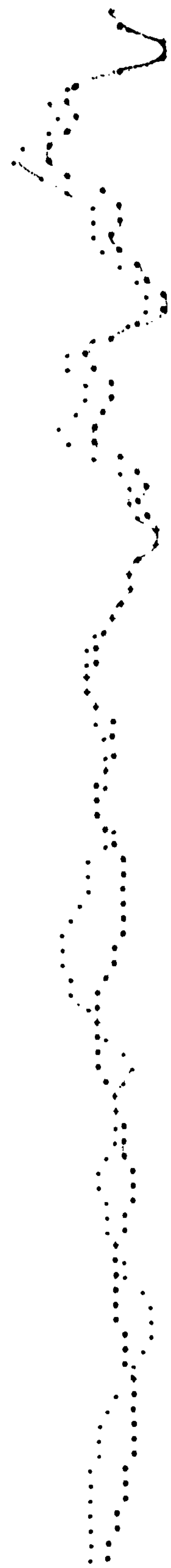
FILE NO. 4 PITCH PERIOD NO. 01 -- TIME OF FIRST PITCH POINT 050 SECONDS, 344 MILLISECONDS



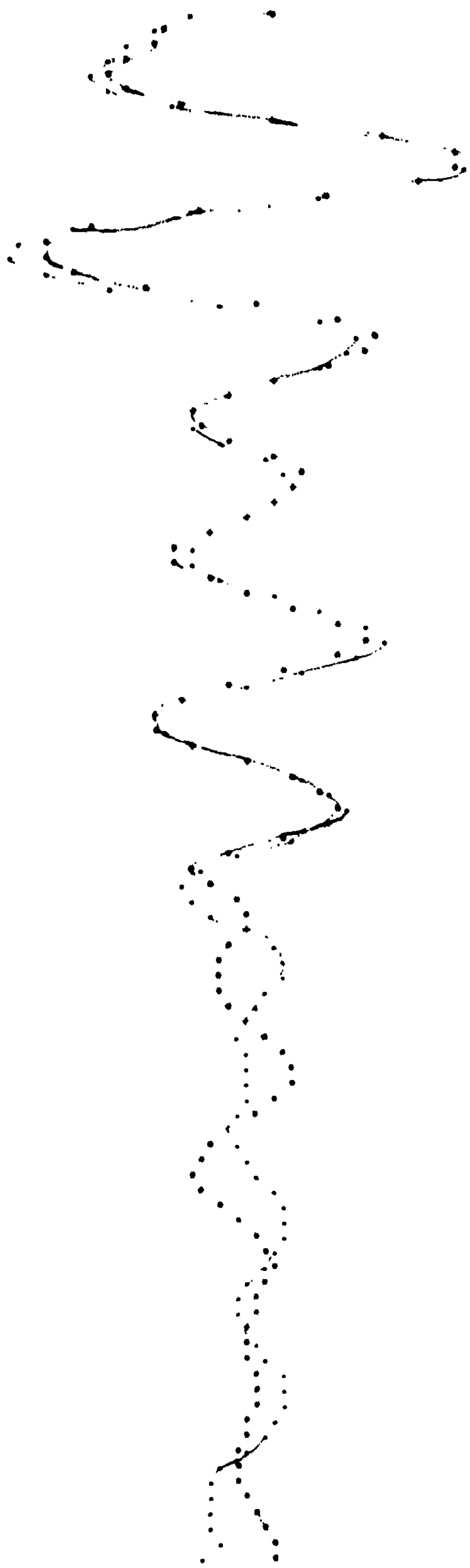
FILE NO. 4 PITCH PERIOD NO. 02 -- TIME OF FIRST PITCH POINT 053 SECONDS, 211 MILLISECONDS



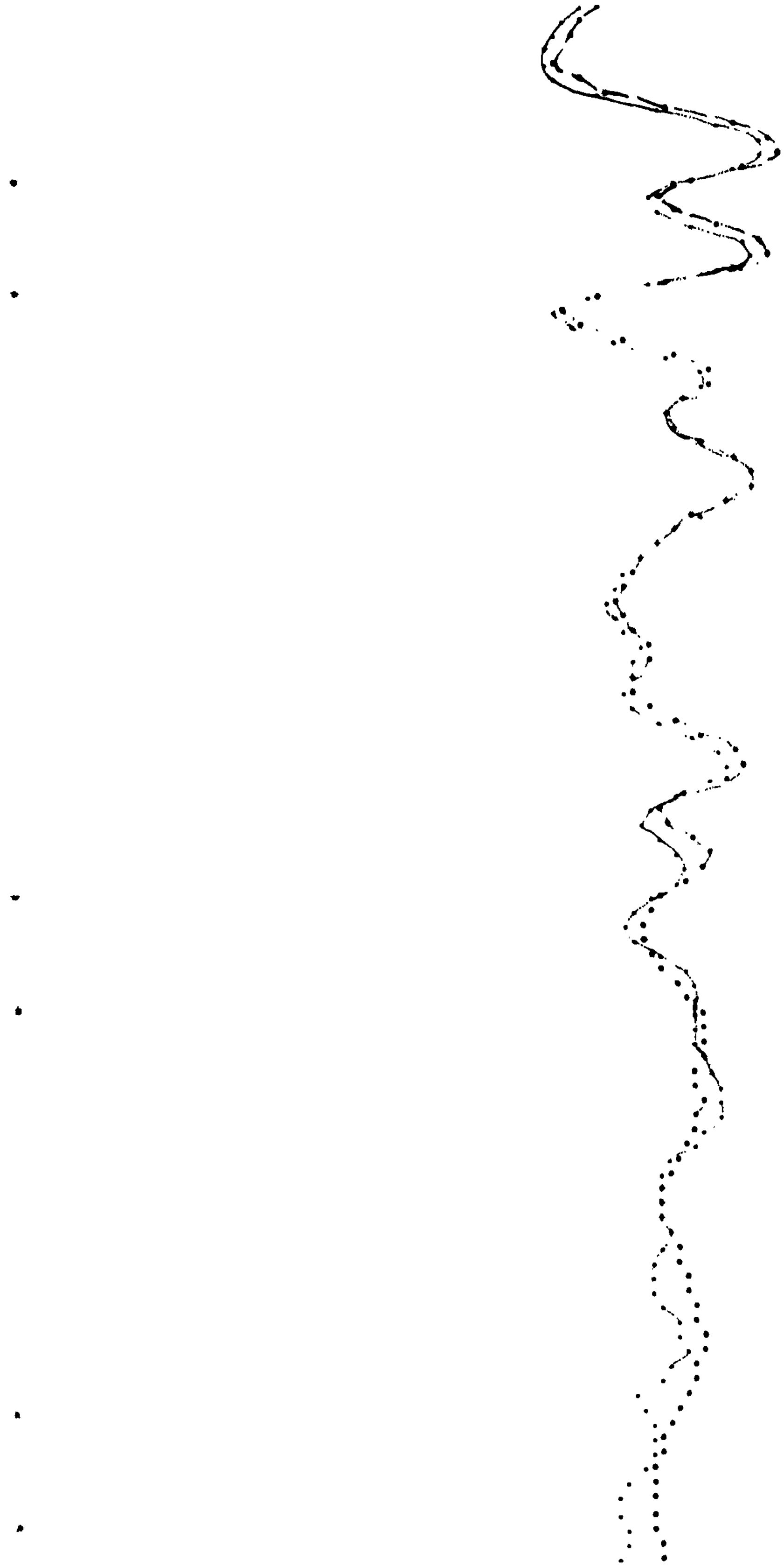
FILE NO. 4 PITCH PERIOD NO. 03 -- TIME OF FIRST PITCH POINT 056 SECONDS, 077 MILLISECONDS



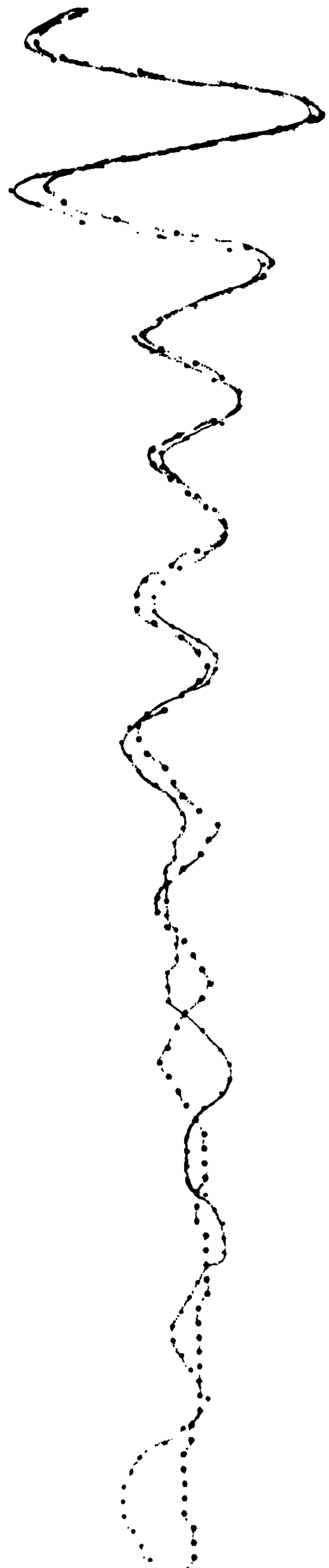
FILE NO. 4 PITCH PERIOD NO. 04 -- TIME OF FIRST PITCH POINT 059 SECONDS, 686 MILLISECONDS



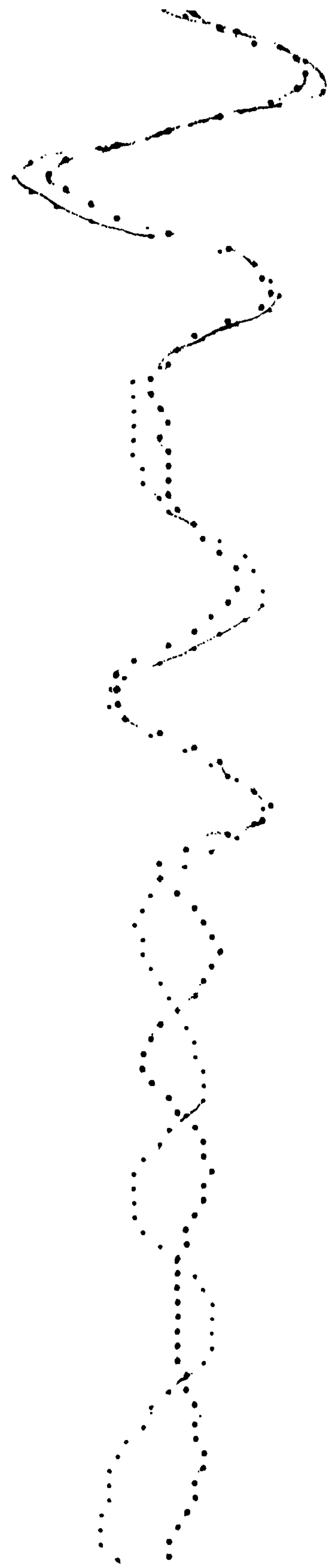
FILE NO. 4 PITCH PERIOD NO. 05 -- TIME OF FIRST PITCH POINT 062 SECONDS, 933 MILLISECONDS



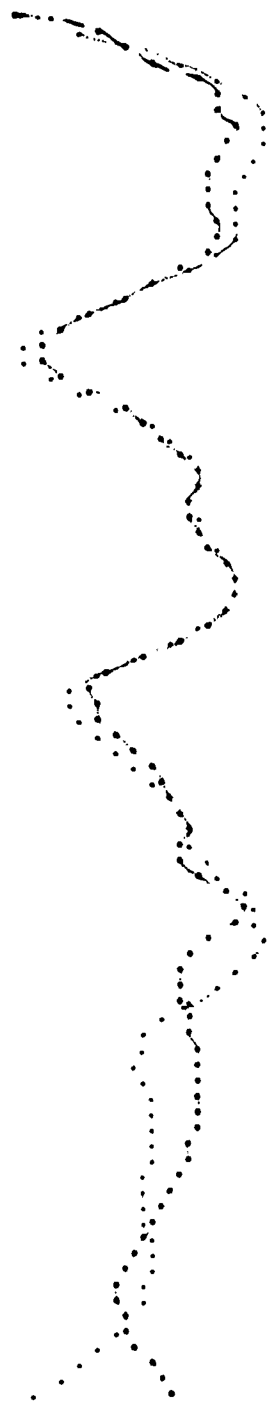
FILE NO. 4 PITCH PERIOD NO. 06 -- TIME OF FIRST PITCH POINT 066 SECONDS, 502 MILLISECONDS



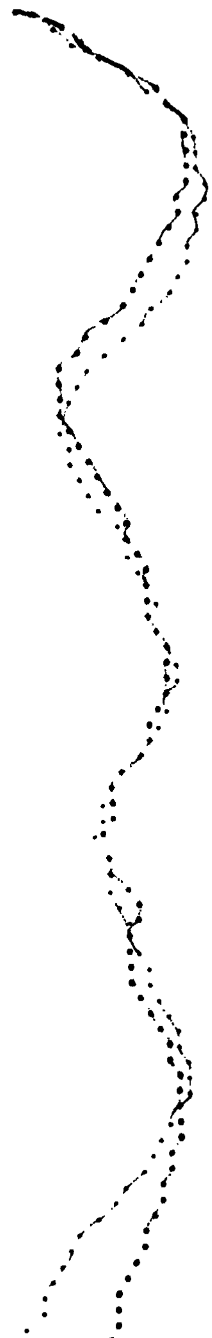
FILE NO. 4 PITCH PERIOD NO. 07 -- TIME OF FIRST PITCH POINT 070 SECONDS, 592 MILLISECONDS



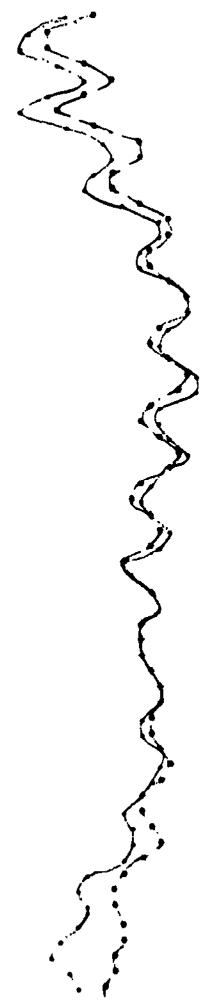
FILE NO. 4 PITCH PERIOD NO. 08 -- TIME OF FIRST PITCH POINT 074 SECONDS, 181 MILLISECONDS



FILE NO. 4 PITCH PERIOD NO. 09 -- TIME OF FIRST PITCH POINT 077 SECONDS, 950 MILLISECONDS

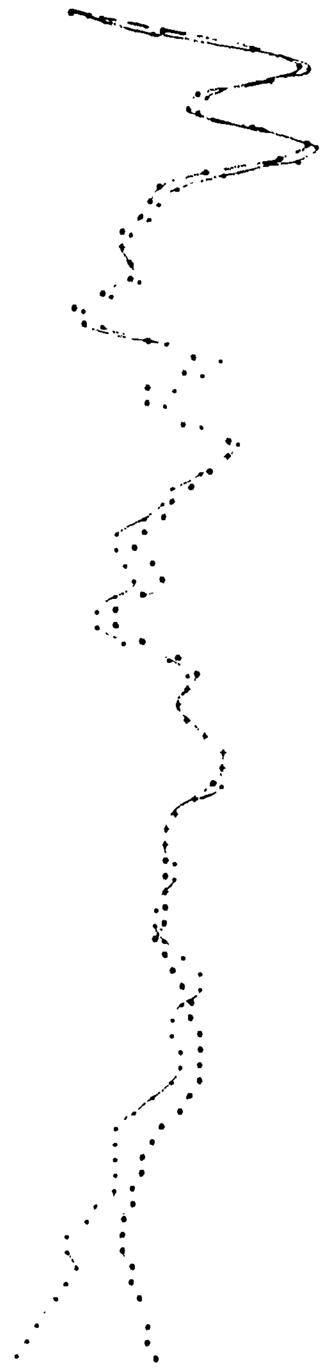


FILE NO. 4 PITCH PERIOD NO. 10 -- TIME OF FIRST PITCH POINT 081 SECONDS, 277 MILLISECONDS

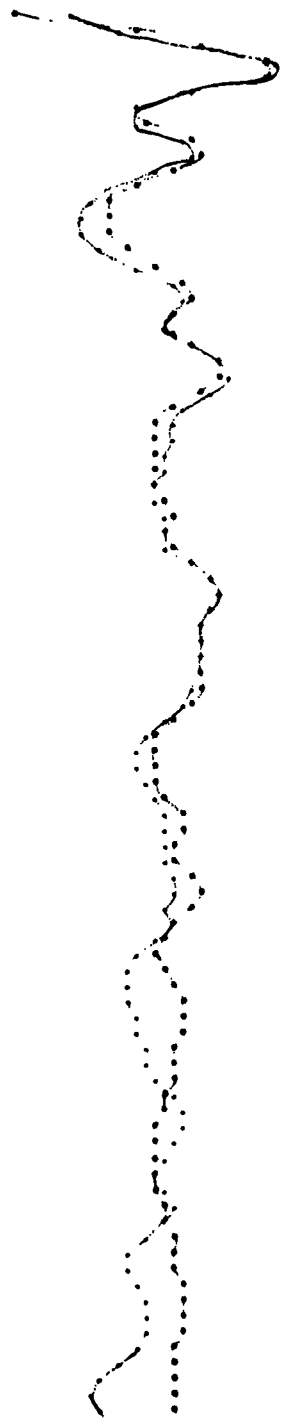


FILE NO. 5 PITCH PERIOD NO. 01 -- TIME OF FIRST PITCH POINT 017 SECONDS, 899 MILLISECONDS

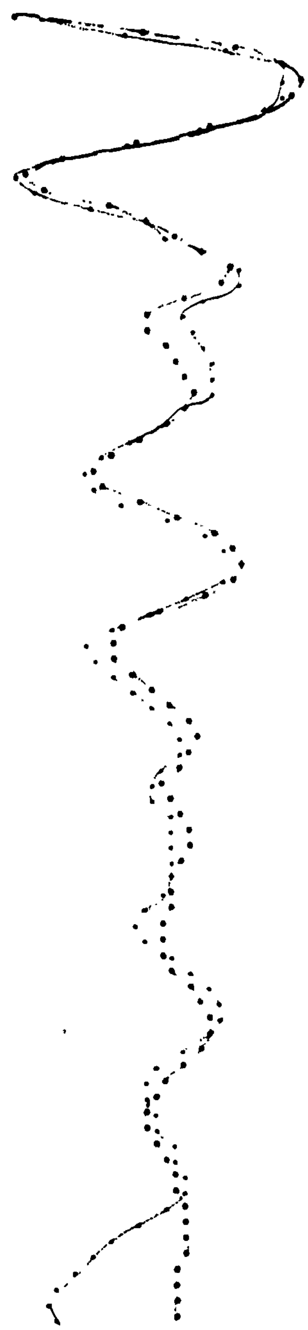
FILE NO. 5 PITCH PERIOD NO. 02 -- TIME OF FIRST PITCH POINT 019 SECONDS, 300 MILLISECONDS



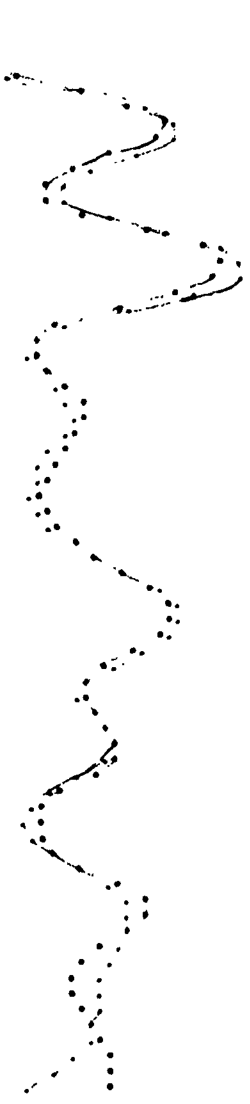
FILE NO. 5 PITCH PERIOD NO. 03 -- TIME OF FIRST PITCH POINT 021 SECONDS, 101 MILLISECONDS



FILE NO. 5 PITCH PERIOD NO. 04 -- TIME OF FIRST PITCH POINT 023 SECONDS, 143 MILLISECONDS

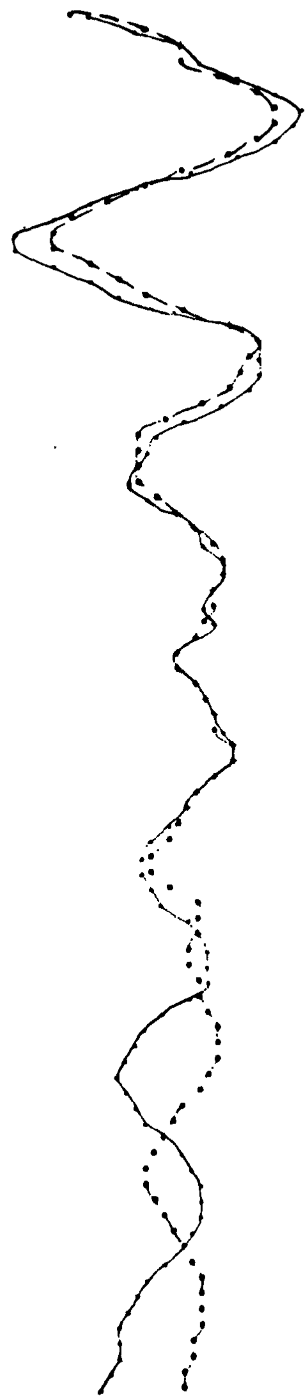


FILE NO. 5 PITCH PERIOD NO. 05 -- TIME OF FIRST PITCH POINT 025 SECONDS, 084 MILLISECONDS



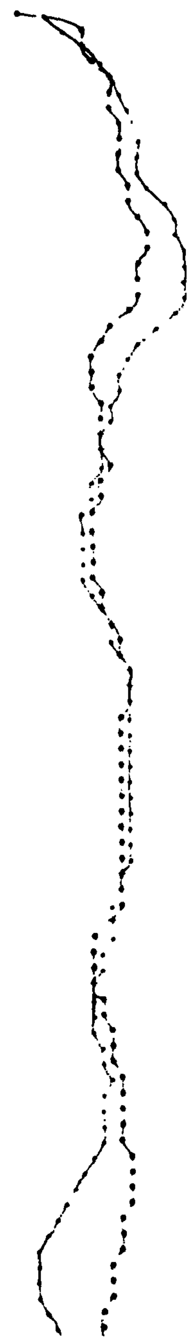
FILE NO. 5 PITCH PERIOD NO. 06 -- TIME OF FIRST PITCH POINT 027 SECONDS, 926 MILLISECONDS

FILE NO. 5 PITCH PERIOD NO. 07 -- TIME OF FIRST PITCH POINT 030 SECONDS, 828 MILLISECONDS



FILE NO. 5 PITCH PERIOD NO. 08 -- TIME OF FIRST PITCH POINT 032 SECONDS, 789 MILLISECONDS

FILE NO. 5 PITCH PERIOD NO. 09 -- TIME OF FIRST PITCH POINT 034 SECONDS, 750 MILLISECONDS



FILE NO. 5 PITCH PERIOD NO. 10 -- TIME OF FIRST PITCH POINT 036 SECONDS, 872 MILLISECONDS

APPENDIX II

A METHOD FOR APPROXIMATING OPTIMUM ORTHOGONAL FUNCTIONS FOR SPEECH ANALYSIS

II.0 INTRODUCTION

One of the goals of our Optimum Speech Program is the determination of an "optimum set of orthonormal functions," where optimum refers to the properties of the associated generalized Fourier-series expansions of speech signals. The desired properties are (a) minimum truncation error (in the mean-square sense) and (b) linearly independent coefficients (in the statistical sense). The Karhunen-Loeve theorem[1] and the work of K.L. Jordan[2] indicate that these properties are provided by the eigenfunctions of the homogeneous Fredholm equation (arranged in order of decreasing eigenvalues) whose kernel is the covariance function characteristic of the speech process (refer to Eq. (II-7)). However, an exact solution of this problem requires the specification of a meaningful covariance function.

To date, insufficient data from which to compute or even determine the existence of a covariance function is available. Indeed, one is hard-pressed to define in advance what would constitute meaningful data, in view of speaker, regional, sixual, age, contextual, syntactic, emotional, etc., variations. Hence, this problem can be solved only in a limited but hopefully non-trivial sense.

The approach taken here utilizes experimentally-determined covariant properties given in terms of the coefficients of an arbitrary Fourier expansion of a speech sample as the data to be used to determine the "optimum" function set -- optimum, at least in the sense discussed in the last paragraph for that particular speech sample used in the computation of the Fourier coefficients. Since explicit expressions for the new functions in terms of the old functions are given (in fact, as linear combinations) the possibilities for and errors involved in truncating the new series are immediately available.

II.1 IMPLICIT SOLUTION

The initial Fourier series expansion is given by

$$x(t) = \underset{N \to \infty}{\text{L.I.M.}} \sum_{n=1}^{N} c_n \phi_n(t), \quad a \leq t \leq b \tag{II-1}$$

where

$$E(C_n C_m) \neq K\,\delta_{nm} \tag{II-2}$$

$$\int_a^b \phi_n(t)\,\phi_m(t)\,dt = \delta_{mn} \tag{II-3}$$

The desired series expansion is

$$x(t) = \underset{M\to\infty}{\text{L.I.M.}} \sum_{m=1}^{M} a_m \lambda_m \psi_m(t), \quad a \le t \le b \tag{II-4}$$

where

$$E(a_m a_n) = \delta_{mn} \tag{II-5}$$

$$\int_a^b \psi_m(t)\psi_n(t)\,dt = \delta_{mn} \tag{II-6}$$

The Karhunen-Loeve integral equation satisfied by the $\psi_m(t)$'s is(1)

$$\lambda_m^2\,\psi_m(t) = \int_a^b K(t,s)\,\psi_m(s)\,ds \tag{II-7}$$

where, according to Mercer's Theorem,(1)

$$K(t,s) = E(x(t)x(s)) = \underset{M\to\infty}{\text{L.I.M.}} \sum_{m=1}^{M} \lambda_m^2 \psi_m(t)\psi_m(s) \tag{II-8}$$

In terms of Eq. (II-1), Eq. (II-8) may also be written as

$$K(t,s) = \underset{N \to \infty}{\text{L.I.M.}} \; E\left(\sum_{n=1}^{N} \sum_{j=1}^{N} C_n C_j \, \phi_n(t)\phi_j(s) \right)$$

$$= \underset{N \to \infty}{\text{L.I.M.}} \; \sum_{n=1}^{N} \sum_{j=1}^{N} \phi_n(t)\phi_j(s) E(C_n C_j) \qquad \text{(II-9)}$$

The last manipulation in Eq. (II-9) must hold if we assume that $x(t)$ is a second-order random process[1] because then all second moments, $E(C_n C_j \phi_n(t)\phi_j(s)) = \phi_n(t)\phi_j(s)E(C_n C_j)$ exist. Consequently, each term on the right-hand side of Eq.(II-9) is equivalent to a summable function in the relevant joint probability measure space and, as a result the summation and integration operations can be interchanged.[3]

Equations (II-9) and (II-9) may be equated (M.S.*) and written as follows:

$$\Psi(t)' \Lambda' \Lambda \Psi(s) \sim \Phi(t)' \, C \, \Phi(s) \qquad \text{(II-10)}$$

It is shown in Addendum A that this is equivalent to Eq. (II-7). Where

$\Psi(t)$ and $\Phi(t)$ are the column matrices whose elements are the ψ_m and ϕ_m respectively,

$\Lambda = \Lambda'$ is the square diagonal M x M matrix whose diagonal elements $\Lambda_{ii} = \lambda_i$,

C is the square N x N, symmetric matrix of elements $E(C_n C_j) = C_{nj}$,

Superscript ' indicates transpose.

If the λ's are equal to the μ's, the square roots of the M eigenvalues of C, then Eq. (II-10) is immediately recognized as the well-known formula for diagonalizing a quadratic relation, where the ψ's are given by

$$\Psi(t) = Q^{-1} \Phi = Q' \Phi \qquad \text{(II-11)}$$

*Indicated by $\sim$.

Q, is of course, the orthogonal matrix whose columns are the eigenvectors of C.

It can be shown that Eq. (II-11) is also a solution of Eq. (II-7) as well as Eq. (II-1(

Thus, by finding the M eigenvalues and eigenvectors of C, a satisfactory set of Ψ's may be determined. However, since we desire an explicit solution to Eq. (II-10) for the expansion functions of Eq. (II-4) in terms of those in Eq. (II-1) in order to be able to evaluate truncation errors, another approach is worthwhile investigating.

II.2 EXPLICIT SOLUTION

The expansion functions in (II-4) are $\left\{\lambda_m \psi_m(t)\right\}$, therefore, if these could be determined as explicit functions of the $\left\{\phi_n(t)\right\}$ and $\left\{C_{nj}\right\}$, then the desired form of the solution is obtained.* This can be done if C in Eq. (II-10) can be represented as the product of a matrix, A, times its transpose A', since then we would have

$$\Lambda \Psi(s) = A' \Phi(s) \tag{II-12}$$

Bodewig[4] and Anderson[5] show us that we can partition the symmetrical matrix C into

$$C = (I + L)\, D (I + L') = (I + U)'\, D (I + U) \tag{II-13}$$

where

L is a lower matrix (only elements below the diagonal are non-zero),

U is an upper matrix (only elements above the diagonal are non-zero),

I is the unit matrix

D is a diagonal matrix with $\det D = \det C = \prod_{i=1}^{N} \mu_i^2$

$(I + U)' = (I + L)$ for a symmetric matrix.

Since D can be partitioned readily into $\sqrt{D}\,\sqrt{D'}$ where $\sqrt{D} = \sqrt{D'}$ is a diagonal matrix which is defined such that its elements are the square roots of the corresponding elements of D, then we can set

---

*The λ_m's themselves are determined later as normalizing constants on the ψ_m's. In this Section the λ_m^2's are not the eigenvalues of C as in Eq. (II-11). The relationship between the two are discussed in Section II.2.

$$A' = \sqrt{D'}\ (I + L') = \sqrt{D'}\ (I + U) \tag{II-14}$$

or

$$A = (I + L)\ \sqrt{D} = (I + U)'\ \sqrt{D}$$

with

$$\det A = \det \sqrt{D} = \prod_{i=1}^{N} a_{ii}\ ; \quad \text{(see Equation (II-12))}$$

Thus, Eq. (II-12) may be written explicitly as

$$\psi_m(t)\ \lambda_m \sim \sum_{n=m}^{n=N} a_{mn}\ \phi_n(t)\ ; \quad m = 1,\cdots,N \tag{II-15}$$

where the a_{mn}'s are the elements of A'.

The upper value, N, on m is imposed by the truncation of the initial series expansion, Eq. (II-1). Hopefully, the series in ψ_m, Eq. (II-4), may be further truncated so that $M < N$. The degree of additional truncation may be determined from the numerical evaluation of Eq. (II-15).

λ_m is determined from Eq. (II-15) and the normalizing relationships Eq. (II-6) and Eq. (II-3)

$$\lambda_m^2 \sim \sum_{n=m}^{N} a_{mn}^2 = N^2(A_{m.})\ ; \quad m = 1,\cdots,M \leq N^* \tag{II-16}$$

---

*Note that

$$\sum_{m=1}^{N} \lambda_m^2 \sim \sum_{m=1}^{N} \sum_{n=m}^{N} a_{mn}^2 = N^2(A') = SP(AA') = Sp(C) = \sum_{n=1}^{N} C_{nn}\ ;$$

where N is the norm and Sp is the spur or trace of the matrix.

Then

$$\psi_m(t) \sim \frac{\sum\limits_{n=m}^{n=m} a_{mn}\, \phi_n(t)}{\left[\sum\limits_{n=m}^{n=N} a_{mn}^2\right]^{1/2}} = \frac{A_m\, \Phi(t)}{N(A_m)}\ ; \quad m = 1,\cdots,M \leq N \tag{II-17}$$

where $A_{m.}$ is a row matrix of elements a_{mn} (with m fixed), N is the norm of a matrix

$$= \sqrt{\sum_i \sum_j a_{ij}^2}$$

Note that Eq. (II-16) results from the orthonormality conditions on the ψ_m's and ϕ_n's. To the author's knowledge, relation Eq. (II-16) does not appear explicitly in the literature, although it utilizes well-known results, since most of the literature refers to matrices without conditions Eq. (II-3) and Eq. (II-6). As a result, Eq. (II-17) appears to be a new, and for our purposes a very useful, result. The author would like to be appraised of the availability of this result elsewhere.

In Addendum B we have derived the following explicit expressions for the a_{mn}'s.

Form I: The a_{mn}'s are the ratios of the following determinants of m x m matrices of appropriate C_{ij}'s.

$$a_{mn} = \frac{a'_{mn}}{\sqrt{a'_{mn}\, a'_{m-1\ m-1}}} = \frac{|C_{11}\ C_{22}\cdots C_{m-1\ m-1}\ C_{mn}|}{\sqrt{|C_{11}\ C_{22}\cdots C_{m-1\ m-1}\ C_{nm}|\ |C_{11}\ C_{22}\cdots C_{m-1\ m-1}|}} \tag{II-18}$$

Where the notation introduced in Eq. (II-18) indicates the determinant of an m x m matrix whose diagonal elements are those shown, i.e,

$$a'_{mn} = \begin{vmatrix} C_{11} & C_{12} & \cdots & C_{1m-1} & C_{1n} \\ C_{21} & C_{22} & \cdots & \cdot & C_{2n} \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot\;\cdot & \cdot & \cdot \\ \cdot & \cdot & & C_{m-1m-1} & \cdot \\ C_{m1} & C_{m2} & \cdots & & C_{mn} \end{vmatrix} \tag{II-19}$$

Only values of a_{mn} for which $n \geq m$ are required.

Form II: An alternative and perhaps computationally more useful set of iterative expressions for the a_{mn}'s are given by the Gaussian or Gauss Doolittle algorithm.[(4),(5)] (In reference (4) many computational extensions and variations to the procedure are also given.)

In this case

$$a_{mn} = \frac{C_{mn}^{(N-1)}}{\sqrt{C_{mm}^{(N-1)}}} \tag{II-20}$$

where $C_{mn}^{(N-1)}$ is an iterated matrix element.

Note that for Eq. (II-17) we don't need the denominators of Eq. (II-18) and (II-20). Thus, if the λ_m's were not required explicitly, we could simplify our computations by just computing the numerators.

In our problem, the C matrix is symmetric, all $C_{ij} \geq 0$; and $C_{ii} \geq C_{ij}$, $i \neq j$.

Note that, from Eq. (II-17), it is seen that

$$\psi_N(t) = \phi_N(t) \tag{II-21}$$

Apparent resemblence to a Gram-Schmidt orthogonalization procedure is to be noted.

We have determined a method for approximating the ψ_n's and λ_n's explicitly in terms of the C_{ij}'s and ϕ_n's; $i,j,n = 1,\cdots,N$. From this a straightforward computation and error analysis can be made.

II.3 SUGGESTIONS, INTERPRETATIONS, AND CONCLUSIONS

It is suggested by this paper that the computations be undertaken, for only by inserting numerical values can the goodness and degree of improvement offered by the new series be determined. The values of the λ_m's (Eq. (II-16)) should decrease with m either in the ordering of the λ_m's suggested by Eq. (II-16) or by reordering. Although the writer did not have time to prove this for the present work, it is believed that the ordering of the λ's can be established before evaluating. The behavior of the a_{mn}'s will indicate the validity of the N-term truncation Eq. (II-16) as an approximation to λ_m. The same may be said for Eq. (II-17) and $\psi_m(t)$. Errors in the new series expansion, Eq. (II-4) are due to two factors: errors in the N-term expansions of λ_m and ψ_m and those caused by truncating M (so that $M < N$). The latter errors (with respect to Eq. (II-1) as a standard) may be computed once we are satisfied with the accuracy of the N-term approximations to the λ_m's and ψ_m's. <u>Since the new functions are linear combinations of the old ones, as given in Eq. (II-17), we are guaranteed of at least as good an approximation using Eq. (II-4) as with Eq. (II-1) for M = N.</u> Thus the N-term approximation to the λ_m's and ψ_m's can result in no worse than as good a representation as the ϕ_n's.

To reiterate a contention made earlier, it is believed that results of Eqs. (II-16) and (II-17) are new. Comments on this point are solicited.

Note that although emphasis has been placed on the explicit solution to Eq. (II-10) given by Eqs. (II-16) and (II-17), the eigenvalue solution, Eq. (II-11) is just as valid a solution. Since the two solutions are not identical and both are valid, it is observed that Eq. (II-10) admits of at least two solutions. The relationships between these are discussed in Appendix III.

ADDENDUM A

EQUIVALENCE BETWEEN SOLUTIONS OF THE KARHUNEN-LOEVE EQUATION AND BILINEAR FORMULA

As an initial step, expand $\psi_m(t)$ in terms of $\phi_n(t)$, and $\phi_n(t)$ in terms of $\psi_m(t)$:

$$\psi_m(t) = \underset{N \to \infty}{\text{L.I.M.}} \sum_{n=1}^{N} b_{mn}\, \phi_n(t) \tag{II-22}$$

$$\phi_n(t) = \underset{M \to \infty}{\text{L.I.M.}} \sum_{m=1}^{M} d_{nm}\, \psi_m(t) \qquad\qquad a \leq t \leq b \tag{II-23}$$

Now

$$b_{mn} = \int_a^b \psi_m(t)\, \phi_m(t)\, dt \tag{II-24a}$$

$$d_{nm} = \int_a^b \phi_m(t)\, \psi_m(t)\, dt \tag{II-24b}$$

therefore

$$b_{mn} = d_{nm} \tag{II-24c}$$

Now apply Eqs. (II-9), (II-22), and (II-3) to (II-8) to obtain a solution to the Karhunen-Loeve equation:

$$\underset{N \to \infty}{\text{L.I.M.}}\ \lambda_m^2 \sum_{n=1}^{N} b_{mn}\, \phi_m(t) = \underset{N \to \infty}{\text{L.I.M.}} \sum_{n=1}^{N} \sum_{j=1}^{N} \phi_n(t) E(C_n C_j) b_{mj} \tag{II-25}$$

Multiply both sides of Eq. (II-25) by Eq. (II-22) and integrate again over t, apply Eq. (II-24) to obtain

$$\lambda_m^2 = \underset{N \to \infty}{\text{L.I.M.}} \sum_{n=1}^{N} \sum_{j=1}^{N} E(C_n C_j)\, b_{mj}\, b_{mn} \tag{II-26}$$

where we have used Eq. (II-27) which is derived from Eqs. (II-5) and (II-22)

$$\underset{N \to \infty}{\text{L.I.M.}} \sum_{n=1}^{N} b_{mm}^2 = 1 \tag{II-27}$$

Consider now the result obtained by equating (II-8) and (II-9); applying Eqs. (II-23) to (II-9) we obtain

$$\underset{N \to \infty}{\text{L.I.M.}} \sum_{m} \lambda_m^2\, \psi_m(t)\, \psi_m(s)$$

$$= \underset{\substack{M \to \infty \\ N \to \infty}}{\text{L.I.M.}} \sum_{j=1}^{N} \sum_{n=1}^{N} \sum_{p=1}^{M} \sum_{q=1}^{M} E(C_n\; C_j) d_{pn}\; d_{qj}\; \psi_p(t)\; \psi_q(s) \tag{II-28}$$

Multiplying each side by $\psi_m(t)\, \psi_m(s)$, integrating over t and s; $[a,b]$, and applying Eqs. (II-6) and (II-24c) provides

$$\lambda_m^2 = \underset{N \to \infty}{\text{L.I.M.}} \sum_{j=1}^{N} \sum_{n=1}^{N} E(C_n\; C_j) b_{mn}\; b_{mj} \tag{II-29}$$

The equivalence of Eqs. (II-26) and (II-29) proves the equivalence of Eqs. (II-7) and (II-10).

ADDENDUM B

DERIVATION OF FORMS FOR a_{mn}

FORM I FOR a_{mn}

Bodewig[6] defines a matrix $Q_i^{(1)}$ according to

$$Q_i^{(1)} = C - x_i y_i'$$

$$Q_i^{(1)} = Q_{i-1}^{(1)} - x_i y_i' = C - \sum_{j=1}^{i} x_j y_j' \tag{II-30}$$

$$\vdots$$

$$Q_N^{(1)} = 0 = Q_{N-1}^{(1)} - x_N y_N' = C - \sum_{j=1}^{N} x_j y_j'$$

so that

$$C = \sum_{j=1}^{N} x_j y_j' \tag{II-31}$$

It can be seen that x_i and y_i' defined as in Eq. (II-32) satisfy these relations

$$x_i = \frac{1}{\sqrt{c_{ii}^{(i-1)}}} \begin{pmatrix} 0 & \cdots & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & \cdots & c_{ii}^{(i=1)} & \cdots & 0 \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & c_{Ni}^{(i=1)} & \cdots & 0 \end{pmatrix} \tag{II-32}$$

i.e., x_i has components

$$\frac{C_{ji}^{(i-1)}}{\sqrt{C_{ii}^{(i-1)}}}, \quad j \geq i$$

where $C_{ji}^{(i-1)}$ are the ji^{th} components of $Q_{i-1}^{(1)}$, $C_{ji}^{(0)} = C_{ji}$. Likewise* y_i' has components

$$\frac{1}{\sqrt{C_{ii}^{(i-1)}}} C_{ij}^{(i-1)}, \quad j \geq i \qquad \text{(II-33)}$$

therefore for symmetrical C, $x_i = y_i$.

It can then be seen that $\sum_{j=1}^{N} x_j$ forms a lower triangle matrix plus diagonal and $\sum_{j=1}^{N} y_j'$ an upper matrix plus diagonal.

$$\left(\sum_{j=1}^{N} x_j y_j' = \sum_{j=1}^{N} x_j \quad \sum_{j=1}^{N} y_j' \right),$$

and therefore,

$$\sum_{j=1}^{N} y_j' = \sqrt{D'}\ (I + U) = A' \qquad \text{(II-34)}$$

Utilizing the formats on pp. 86 ff. of Bodewig (and our notation), it can be seen that

---

*Alternatively x_i and y_i could have been defined as $s_i e_i'$ and $e_i s_{\varphi}'$ according to the notation introduced further on.

$$C_{11}Q_1^{(1)} = B$$

$$C_{22}^{(1)}Q_2^{(1)} = C$$

$$C_{33}^{(2)}Q_3^{(1)} = D \tag{II-35}$$

$$C_{44}^{(3)}Q_{44}^{(1)} = E$$

$$\begin{array}{cc} \cdot & \cdot\ \cdot \\ \cdot & \cdot\ \cdot \\ \cdot & \cdot\ \cdot \end{array}$$

Then, the $C_{ij}^{(p)}$'s become

$$C_{ij}^{(p)} = \frac{|C_{11}\ C_{22}\ \cdots\ C_{pp}\ C_{ij}|}{|C_{11}\ C_{22}\ \cdots\ C_{pp}|} \tag{II-36}$$

and the elements of A', a_{ij}, are

$$\frac{C_{ij}^{(i-1)}}{C_{ii}^{(i-1)}} = \frac{|C_{11}\ C_{22}\ \cdots\ C_{i-1\ i-1}\ C_{ij}|}{\sqrt{|C_{11}\ C_{22}\ \cdots\ C_{i-1\ i-1}|\ |C_{11}\ C_{22}\ \cdots\ C_{ii}|}}$$

FORM II FOR a_{mn}

According to Bodewig[7]

$$D + DU = Q_{N-1}^{(2)} \cdots Q_2^{(2)}Q_1^{(2)}\ C \tag{II-37}$$

where D and U were defined previously (in Eq. (II-13) and what follows it) and $Q_i^{(2)}$ is defined by

$$Q_i^{(2)} = I + s_i e_i' \tag{II-38}$$

s_i is a column vector all of whose elements above and including the i^{th} one vanish and e_i' is a unit row vector with only the i^{th} element occupied, i.e.,

$$e_1' = (1,0,0,0,\cdots,0)$$

$$e_2' = (0,1,0,0,\cdots,0)$$

$$\begin{matrix} \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \end{matrix}$$

Therefore, $s_i e_i'$ is an N x N matrix with zero elements everywhere but the $i + 1^{th}$ through N^{th} elements of the i^{th} column.

The s_i are explicitly column vectors whose elements are

$$-\frac{C_{pi}^{(i-1)}}{C_{ii}^{(i-1)}}, \quad p = i + 1,\cdots,N \quad . \tag{II-39}$$

where here the $C_{pi}^{(i-1)}$ are the pi^{th} elements of $Q_{i-1}^{(2)} \cdots Q_1^{(2)} C$ with $C_{pi}^0 \equiv C_{pi}$. Specifically this method gives us an iterative procedure for finding D + CU.

Since the elements of D are $C_{ii}^{(N-1)}$, then, according to Eq. (II-14) in the body of this paper, the a_{mn}'s are given by

$$a_{mn} = \frac{C_{mn}^{(N-1)}}{\sqrt{C_{mm}^{(N-1)}}} \tag{II-40}$$

and our iterative procedure can provide a_{mn} directly. Also, in our particular problem, from (II-40),

$$\det D = \prod_{i=1}^{N} C_{ii}^{(N-1)} = \prod_{i=1}^{N} a_{ii}^2 \tag{II-41}$$

ADDENDUM C

RELATIONSHIP BETWEEN TWO METHODS OF DIAGONALIZATION

A symmetric matric C may be represented as the product of two matrices where one is the transpose of the other in an infinite number of ways, all related by orthogonal matrices. Thus, if A is determined by the method of Section II.2, then the eigenvalues of C, the u^2's, may be determined from the eigenvalue Eq. (II-42) as well as (II-45).

$$PA'\,x = u\,x \tag{II-42}$$

where P is an orthogonal matrix chosen to make Eq. (II-42) hold. The u's themselves are determined by the secular determinant

$$|PA' - u| = 0 \tag{II-43}$$

and

$$AP'\,PA' = AA' = C \tag{II-44}$$

Recalling that the relations for C are

$$Cx = u^2x \tag{II-45}$$

$$|C - u^2| = 0 \tag{II-46}$$

where x and u are the same as in Eqs. (II-42) and (II-43).*

The eigenvalues of the triangular plus diagonal matrix A' are η, given

$$|A' - \eta| = 0 \tag{II-47}$$

---

*The necessary and sufficient condition for this is that PA' and AP' commute.[8] (It follows then that these also commute with C.)

or

$$\prod_{i=1}^{N} (a_{ii} - \eta) = 0$$

In fact, then, $\eta_i = a_{ii}$ and the eigenvalue equation for A becomes

$$A'y = \eta \, y \tag{II-48}$$

or

$$y'A = y \, \eta$$

Now, we cannot write,

$$Cy = \eta^2 y$$

since A and A', as determined in Section (II.2) do not commute, i.e.,

$$A'A \neq A \, A'^{(8)} \quad .$$

However, we can obtain by premultiplying both scales of Eq. (II-48) by P.

$$ux = P\eta y \tag{II-49}$$

and multiplying each side by its transpose.

$$u^2 x'x = \eta^2 y'y \tag{II-50}$$

Since the y's are not necessarily orthogonal we cannot write

$$X \, M^2 X' = Y \, N^2 Y' \tag{II-51}$$

where X and Y are matrices whose columns are the eigenvectors x and y, respectively, and M^2 and N^2 are the diagonal matrices whose elements are the u^2 and η^2 terms. In fact, as we know

$$XM^2X' = C = A\,A' = (Y^{-1})'\; NY'Y\,N\,Y^{-1}$$

$$= (Y')^{-1}\; NY'Y\,N\,Y^{-1} \tag{II-52}$$

where, although

$$Y' \neq Y^{-1}\;,\; (X^{-1})' = (Y')^{-1}\;.$$

Note that since A and A' do not commute

$$Y\,N\,Y^{-1}(Y^{-1})'NY' \neq C \tag{II-53}$$

Consequently, the eigenfunctions of A or A', $\left\{\eta_i\right\}$, are not useful for obtaining a diagonal form.

However, it is useful to point out the following relationships:[9]

$$\det C = \det AA' = (\det \sqrt{D})^2$$

$$= \det D = \prod_{i=1}^{N} u_i^2 = \prod_{i=1}^{N} a_{ii}^2 = \prod_{i=1}^{N} \eta_i^2 \neq \prod_{i=1}^{N} \lambda_i^2 \tag{II-54}$$

where λ_i are the constants determined in Eq.(II-17), and

$$Sp(c) = Sp(AA') = Sp(A'A) = N^2(A') = \sum_{m=1}^{N}\sum_{n=m}^{N} a_{mn}^2$$

$$= \sum_{n=1}^{N} C_{nn} = \sum_{i=1}^{N} u_i^2 = \sum_{i=1}^{N} \eta_i^2 + \sum_{m=1}^{N}\sum_{n=m+1}^{N} a_{mn}^2$$

$$\sim \sum_{m=1}^{N} \lambda_m^2 \geq \sum_{i=1}^{N} \eta_i^2 = (Sp\;A)^2 = (Sp\;A')^2 = Sp^2(A) = Sp^2(A') \tag{II-55}$$

Now if we consider as a measure of the goodness of the finite series fit, the time-average of the mean square difference between the complete series and the truncated series $\langle \bar{e}^2 \rangle$ then $\langle \bar{e}^2 \rangle$ for the truncated form of the sense whose terms are given by Eqs. (II-16) and (II-17) is

$$\langle \bar{e}^2 \rangle = \sum_{m=M+1}^{\infty} \lambda_m^2 = \langle K(t,t) \rangle - \sum_{m=1}^{M} \lambda_m^2$$

$$= \langle K(t,t) \rangle - \sum_{n=1}^{N} C_{nn} + \sum_{m=M+1}^{N} \lambda_m^2 \quad . \tag{II-56}$$

For the eigenfunction series we have for an M term series

$$\langle \bar{e}^2 \rangle = \langle K(t,t) \rangle - \sum_{n=1}^{N} C_{nn} + \sum_{i=M+1}^{N} u_i^2 \tag{II-57}$$

For the original series (II-1), we have

$$\langle \bar{e}^2 \rangle = \langle K(t,t) \rangle - \sum_{n=1}^{N} C_{nn} + \sum_{j=M+1}^{N} C_{jj} \tag{II-58}$$

(It is important to remember that the λ's and u's are computed from the N x N matrix of C's. We then truncate to M < N terms. If they were computed from the corresponding M x M matrix of C's, they would all have the same error as noted below for M = N.)

Consequently, the relative goodness of each series in truncation is measured by the smallness of the respective terms.

$$\sum_{m=M+1}^{N} \lambda_m^2 \, , \quad \sum_{i=M+1}^{N} u_i^2 \, , \quad \sum_{j=M+1}^{N} C_{jj} \tag{II-59}$$

where we note that for $N = M$ they are each of equal goodness.

Now, since a_{mn} is an element of A' and $C = AA'$, it can be seen that

$$C_{ij} = \sum_{m=1}^{\text{longer of } j \text{ or } i} a_{mj}\, a_{mi}\ , \quad \text{and} \quad C_{nn} = \sum_{m=1}^{n} a_{mn}^2\ .$$

Then,

$$\sum_{n=M+1}^{N} C_{nn} = \sum_{n=M+1}^{N} \sum_{m=1}^{n} a_{mn}^2 = \sum_{n=m+1}^{N} \sum_{m=1}^{M+1} a_{mn}^2 + \sum_{n=m+1}^{N} \sum_{n=m}^{N} a_{mn}^2$$

$$= \sum_{n=M+1}^{N} \sum_{m=1}^{M+1} a_{mn}^2 + \sum_{m=M+1}^{N} \lambda_m^2 \geq \sum_{m=M+1}^{N} \lambda_m^2\ . \qquad \text{(II-60)}$$

A possible decrease in time-average, mean-truncation error is gained over the ϕ's by the ψ's of Eq. (II-17). However, if the value $\bar{e}^2$ (not time-averaged but the point-by-point mean error is used), as a criterion an even greater decrease is possible due to the appearance of cross products in the ϕ expression. To wit:

$$\bar{e}_\phi^2 = K(t,t) - \sum_{n=1}^{M} \sum_{j=1}^{M} C_{nj}\, \phi_n(t)\, \phi_j(t) - 2 \sum_{n=1}^{M} \sum_{j=M+1}^{N} C_{nj}\, \phi_n(t)\, \phi_j(t) \qquad \text{(II-61)}$$

$$\bar{e}_\psi^2 = K(t,t) - \sum_{n=1}^{N} C_{nn}\, \psi_n^2(t) + \sum_{i=M+1}^{N} \lambda_m^2\, \psi_m^2(t) \qquad \text{(II-62)}$$

In order to compare the u expression, with more than the following mode approximation, it appears that more has to be known about the eigenfunction, i.e.,

$$\sum_{i=M+1}^{N} u_i^2 = \sum_{i=M+1}^{N} \sum_{p=1}^{N} \sum_{q=1}^{N} C_{pq}\, x_p^{(i)}\, x_q^{(i)} \tag{II-63}$$

Since $C_{pp} \geq C_{pq}$, $p \neq q$, then

$$\sum_{i=M+1}^{N} u_i^2 \leq \sum_{i=M+1}^{N} \left(\sum_{p=1}^{N} \sqrt{c_{pp}}\, x_p^{(i)} \right)^2 \tag{II-64}$$

REFERENCES

1. Loeve, M., Probability Theory, Van Nostrand, N.Y., 1960, p. 478.
2. Jordan, K., M.I.T., R.L.E., Q.P.R., No. 52, 1959, p. 70.
3. Kestelman, H., Modern Theories of Integration, Dover, N.Y., 1960, p. 14.
4. Bodewig, E., Matrix Calculus, Interscience, N.Y., 1956, pp. 17f, pp. 86f, pp. 110.
5. Anderson, T.W., Introduction to Multivariate Statistical Analysis, Wiley, N.Y., 1960, p. 185, pp. 349ff.
6. Bodewig, E., Matrix Calculus, Interscience, N.Y., 1956, pp. 17ff.
7. Bodewig, E., Matrix Calculus, Interscience, N.Y., 1956, pp. 101 and 110.

APPENDIX III

PROGRAM IDENTIFICATION - SE002

I. PURPOSE

To extract characteristic pitch periods, analyze and reconstruct them, and "plot" the original and reconstructed points.

II. METHOD

After the orthonormal functions have been read, the program begins reading the input tape. From each record a pitch period is extracted, the bounds of which are specified on parameter cards. (For the same reasons as in the previous program, SE001, the parameters were assembled with the program.) The pitch points are individually converted to floating point and this new form of the pitch period is analyzed using the orthonormal functions such that thirty-two coefficients result. Besides being written out with two identification words as a separate output, these coefficients are used to reconstruct the speech.

As each point in the pitch period is reconstructed, the original and the reconstructed points are "plotted" on a scale of ten to one in a BCD output area. The original points are plotted as periods, ".", and the reconstructed points as asterisks, "*". If a reconstructed point exceeds 1023 or is negative, an "x" is plotted in the highest or lowest plot position respectively. A "+" indicates that the scaled original and reconstructed points are plotted in the same position.

When twenty pitch periods in each of five files have been handled, the program prints a message on-line and stops.

III. DATA REQUIREMENTS

A. Parameters

1. Four parameters are required for each pitch period.
2. The first parameter appears as a "BCI" card containing the time in BCD, three characters for seconds and three for milliseconds. These must match the time on the portion of the input record in which the pitch period begins.

3. The second parameter indicates the actual starting point of the pitch period. It takes the form of a TXI instruction with a fixed address, "XEC01", fixed tag, "4", and a negative variable decrement which is the actual number of the beginning point.

4. The third parameter is similar to the first except that it contains the time of the portion of the input record in which the pitch period ends.

5. The fourth parameter is similar to the second but it defines the end point of the pitch period. It takes the form of a TXI instruction with a fixed address, "XEC02", fixed tag, "4", and a positive variable decrement which is the actual number of the end point of the pitch period.

6. The program is written to handle five files of twenty pitch periods each, hence there are five sets of eighty parameters each or a total of four hundred.

7. As in SE001, the parameters were assembled with the symbolic deck in the interest of speed.

8. Changes to any one parameter or set of same, is discussed under the modifications' section of this report.

B. INPUT

1. Orthonormal Functions Tape

    a. This input is on a low density magnetic tape, created in binary mode.
    b. There is one file of data.
    c. The file consists of thirty-two 301-word records.
    d. Each function is represented in floating point binary.

2. Pitch Point Tape

    a. The input data is on a high density magnetic tape, created in binary mode by the program identified as SE001.
    b. There are five files of data on the tape.
    c. Each file contains twenty 305-word records.
    d. Each 305-word record can be subdivided into five 61-word subrecords.

e. Each subrecord has as its first word a BCD representation of time, three characters for seconds and three for milliseconds.

f. Each of the remaining sixty words contains a pitch point in binary, which range from zero (0) to one thousand and twenty-three (1,023).

g. The pitch points are right justified.

C. OUTPUT

1. Coefficients

a. The coefficients are written on a high density magnetic tape.

b. There are five files each containing twenty 34-word records.

c. The first two words of the thirty-four are identification words. Word one contains the file number in the decrement portion, and the number of points in the analyzed pitch period in the address portion. Word two contains the BCD representation of the time of the record in which the pitch period began.

d. Words three through thirty-four contain thirty-two coefficients in normalized floating point.

2. "PLOT"

a. The "plot" is on a high density magnetic tape, written in BCD mode.

b. It contains one file of blocked records, five 22-word records per block.

c. There is a header record for each pitch period. It contains the file and pitch period numbers, as well as the time of the input subrecord in which the pitch period begins.

d. The header record is followed by 32 coefficient records. The odd numbered coefficients have in the same record the square root of the sum of the squares of the odd numbered coefficient and the next even one.

e. The coefficients are followed by as many records as there are points in the pitch period. The format for these records is as follows:

Position Number	Contents
1 - 3	pitch point number
7 - 10	original pitch point
13 - 16	reconstructed pitch point
25 - 127	the "plot" of the original and reconstructed pitch points, "." for the original, "*" for reconstructed, "+" if both are in the same position, "X" if off scale.

3. Unreadable Records

Any record which cannot be read is reread 10 times. If the reading still fails the record is written, as is, on the unreadable records tape. This operation is a fixed function of the I/O package.

IV. MACROS

All macros used are part of the Input-Output package written by A/2c Robert Barger at the Air Force Intelligence Center, Washington, D. C. These include:

QFILE - a macro to define the files and set certain options in the program.

QOPEN - a macro to initialize the I/O system.

QWRITE - a macro to write logical records on tape.

QFINSH - a macro to finish writing a block initiated by QWRITE.

and QCLOSE - a macro to wind up all I/O functions.

V. SUBROUTINES

A. IORDWR - a subroutine in the I/O package to read or write tape depending on the calling sequence.

B. BINDEC - a subroutine developed at AFIC to convert binary to BCD.

C. SQRT - a subroutine currently available as part of FMS to take square roots.

VI. MACHINE CONFIGURATION

A. 32K 7090 with two channels A and B

B. 5 tape drives on Channel A

C. 4 tape drives on Channel B

D. 1 card reader

E. 1 printer

VII. OPERATING INSTRUCTIONS

A. Load IB Monitor (SOS) system tape on A1

B. Load orthonormal functions tape on B4

C. Load pitch point data tape on A6

D. Load blanks on:

A2 - required by the system

A4 - binary output - coefficients tape

A7 - unreadable records tape

B1 - required by the system

B2 - required by the system

B6 - BCD output - "plot" tape.

E. Put squoze deck in card reader hopper and ready.

F. Put sense switch 1 on, all others off.

G. Clear memory.

H. Load tape.

I. At END-OF-JOB

1. return system tape (A1), and orthonormal functions tape to programmer.
2. return binary input (A6) to storage (optional)
3. label and store binary output (A4)
4. if no input records were unreadable, release A7 immediately. If there were unreadable records, releasing is at the option of the programmer.
5. list BCD "plot" tape (B6) on 1401 with special program which prints blocked records.

VIII. ERROR MESSAGE

	Message on Line	Cause	Corrective Action
A.	"ILLEGAL EOF ON ORTHONORMAL FUNCTTIONS TAPE."	stated in message	1.) This stop should never happen. 2.) Check mod package to make sure nothing is interfering with this macro. 3.) If macro is O.K., regenerate the tape and rerun.
B.	"ILLEGAL EOF ON INPUT TAPE."	stated	1.) The input tape is required to have five files, each containing 20 305-word records. 2.) Check the input tape and regenerate if necessary.
C.	"INPUT TIME HIGHER THAN SELECTION TIME."	The time on the input record is higher than the selection time specified for the beginning of the selection.	1.) Check the mod package to make sure there are no overlapping alter cards. 2.) Check for keypunch errors in the parameters. If so, correct and rerun.

Message on Line	Cause	Corrective Action
		3.) Check listing of previous program to make sure the beginning and end points are actually times of the records on the input data tape. Correct and rerun.
D. "NO INPUT TIME MATCHES CURRENT SELECTION TIME."	There is no match in the input record for the selection time specified.	Same as for C.
E. "IN TIME HIGHER THAN END SELECTION TIME."	The time on the input record is higher than the time specified for the end of the selection.	Same as for C.
F. "TROUBLE ADVANCING TO NEXT FILE. ILLEGAL EOF."	An extra end of file is detected somewhere between files.	1.) Check for multiple tape marks between files on the pitch period tape. 2.) Regenerate if necessary.

N.B. If necessary, a system Dump may be produced by manually transferring to 112_{10} or 160_8 .

IX. MODIFICATIONS

Modifications may be made to any single parameter in one of the following ways:

A. Changing the "beginning" time:

Find in the listing the "beginning time" to be changed. [Symbolic address "BANDE" + 80 (File number -1) + 4 (Selection number -1)] . This will appear as BCI 1,xxxxxx, with x's equal to the time in seconds and milliseconds. Find the alter number associated with it. This is in the column to the left of the symbolic address field in the listing. Punch the alter number twice in succession separated by a comma. Punch a new BCI card replacing the old time with the new. Put the change in mod package and run.

e.g. To change the beginning time of third selection of file 2:

BANDE + 80 (2 - 1) + 4(3 - 1) = BANDE + 88 or alter number 1835.

Punch

Column 8	16
ALTER	1835,1835
BCI	1,YYYYYY

where the Y's represent the new time.

B. Changing the number of the first point.

Find in the listing the point to be changed. ["BANDE" + 80 File number - 1) + 4(Selection number - 1) + 1] . It will appear as a "TXI XEC01, 4, -X." where "X" is the point number. Find the associated alter number. Punch the number in an alter card. Punch a new TXI card replacing the old point number with the new. Put the two cards in mod package and run.

e.g. To change the beginning point number of the third selection of file 2:

BANDE + 80 (2 - 1) + 4(3 - 1) + 1 = BANDE + 89 or alter number 1836.

Punch

Column 8	16
ALTER	1836,1836
TXI	XEC01,4,-Y

where Y represents the new point number.

C. Changing the end time is the same as changing the beginning time, but add 2 to the alter number.

e.g. To change the end time of the third selection of file 2:

Punch

Column 8	16
ALTER	1837,1837
BCI	1,YYYYYY

where Y's represent the new time in seconds and milliseconds.

D. Changing the number of the end point is analogous to changing the number of the first point except the alter number is 2 more than the one for the beginning point, the address is XEC02, and the decrement is positive.

e.g. To change the end point number of the third selection of file 2:

Punch

```
Column 8        16
       ALTER    1838,1838
       TXI      XEC02,4,Y
```

where Y represents the new point number.

E. If the set of four parameters for a pitch period is to be changed, the individual parameters would be punched according to the format above. The four alter cards would be replaced by one, punched as follows for the preceding example.

```
Column 8        16
       ALTER    1835,1838
```

APPENDIX IV

PROGRAM IDENTIFICATION - SE001

I. PURPOSE

To select sets of records from five sound streams, each set containing one or more representative pitch periods, and to plot the pitch points contained therein.

II. METHOD

Since the program is tailor made for a particular input, a brief description of this input is in order. Five people were recorded, each one saying ten sounds and then repeating them. The sounds were digitized and the resultant record placed on magnetic tape. This tape is the program input. The parameter - - the numbers of the records in which the sponsor was interested - - were assembled with the original symbolic deck in order to avoid the necessity of reading a parameter tape.

Before a record is read, the record counter is checked to see if it is of any interest. If it is not, the counter is incremented by one, the record effectively skipped, and the next record considered. If it is a selected record, a header consisting of file number, selection number, and time, in seconds and milliseconds, is written on the BCD output tape. The time in BCD is also stored in the first word of the binary output area.

Each point contained in the record is treated separately. First, the point is unpacked from the original format of 3 points/word. If, when tested, it is found to be greater than 1023, it is replaced by 1023, stored in the binary output area, and a flag set to indicate this condition on the "plot." Otherwise, the input number is stored in the binary output area and then examined to determine its "plot" position. The "plot" is in reality a block of 103 positions of the BCD output area in which a character, a period, is stored depending on the magnitude of the given number after it has been scaled ten to one. After it has been plotted, the number and the running pitch count are converted to BCD and both are stored in the output BCD record and the record written. This process continues until the 20 words, 60 points, of input are completed. Four more input records, in sequence, are added to this group before the selection search is repeated. When 20 selections for the first file are processed, the input is spaced forward to commence similar treatment of the remaining four files. When all the data is completed, the output is finished and a message is printed on line.

III. DATA REQUIREMENTS

A. Parameters:

The parameters in this program are a little unusual in that they are actual instructions assembled with the program. The input records were counted, beginning with zero, and the counts associated with the first record of the sets of five for each selection were supplied to the program as decrements of TXL instructions. This approach was used to avoid the necessity of reading a parameter tape (the I O package being used excludes the possibility of reading parameters from cards) and to decrease the execution time of the program. The section on modifications in this write-up will illustrate how modifications can be made.

B. INPUT

1. The input data is on a low density, binary mode, magnetic tape.

2. There must be five files of data on the tape.

3. Each record is 21 words long. The first word is in three parts. The first part, 12 bits, is an ID not used in this program. The next two parts are time specifications; 12 for time in seconds and 12 for milliseconds. The remaining 20 words contain pitch points of 12 bits each, packed three per word.

4. Of the 12 bit pitch point, bit zero is unused. Bit 1 indicates an overflow condition if it is one. The remaining 10 bits can contain a number equal to or less than 1777 in octal or 1023 in decimal.

C. OUTPUT

1. Binary Output

   a) The binary output is a magnetic tape written in high density.
   b) It contains five files.
   c) Each file contains 20 305 - word records.
   d) Each record is the expansion of five consecutive input records. The first word is the time of the first record the set in BCD, three characters for seconds, and three for milliseconds. The next 60 words are the unpacked pitch points of the first record, with overflow conditions removed. The format is repeated for each of the other four records in the set.

2. BCD OUTPUT

a) The BCD output is a high density magnetic tape.
b) There is one file of blocked records, 5 22-word records/block.
c) A major header record is written once for each selection. It contains the file number, one through five, the selection number, one through twenty, and the number of the input record.
d) A minor header is written for each input record in the selection; i.e. five per selection. It contains the time in seconds and milliseconds of the individual record.
e) Every minor header is followed by 60 point records. Each of these contains the number of the pitch point, the pitch point, and the "plot" of the pitch point. The number of the pitch point occupies character positions one and two of the record. The pitch point itself occupies position seven through ten. The "plot" character can appear anywhere from position 25 through 127.
f) When an input file is completed, a record of asterisks is written on the BCD output.

3. UNREADABLE RECORDS

a) A record is not deemed unreadable until it has been read and reread ten times.
b) An unreadable record is written on A7 (fixed by the IO package) in the same form as it is read.

IV. MACROS

All the macros used are part of the Input-Output package written by A/2c Robert Barger at the Air Force Intelligence Center, Washington, D.C. These include:

QFILE - a macro to define the files and set certain options in the program.

QOPEN - a macro to initialize the I - O system.

QWRITE - a macro to write logical records on tape.

QFINSH - a macro to finish a writing block initiated by QWRITE, and

QCLOSE - a macro to wind up all I - O functions.

V. SUBROUTINES

A. IORDWR - a subroutine in the IO Package to read or write tape depending on the calling sequence.

B. BINDEC - a subroutine, developed at AFIC, to convert from binary to decimal.

VI. MACHINE CONFIGURATION

A. 32K 7090 with 2 channels, A and B.

B. 5 tape drives on channel A

C. 3 tape drives on channel B

D. 1 card reader

E. 1 printer

VII. OPERATING INSTRUCTIONS

A. Load IB MONITER SOS System tape on A1

B. Load input tape on A4

C. Load blanks on:

A2 - required by the system

A6 - binary output tape

A7 - unreadable records tape

B1 - required by the system

B2 - required by the system

B4 - BCD output

D. Put squoze deck in card reader hopper and ready.

E. Put sense switch I on, all others off.

F. Clear memory.

G. Load tape.

H. At END-OF-JOB:

1) Return A4 to A1 to programmer.
2) List B4 with special 1401 print program which prints blocked records.
3) Label and save A6.
4) Saving A7 is optional.

VIII. ERROR MESSAGE

A. "THERE IS AN ILLEGAL EOF" will be printed if an unexpected EOF occurs. The final program halt will indicate at which point in the progran it occurred.

HTR 10226_8 - when a record of interest is being read.

HTR 10227_8 - when a record of no interest is being skipped.

HRT 10230_8 - When the tape is being spaced forward after a legitimate EOF

B. A dump may be taken by manually transferring to 10231_8 , or if this fails, by transferring to 160_8 (112_{10}).

IX. MODIFICATIONS

If for any reason the record number of the first record of a selected set must be changed, this can be accomplished easily by the following procedure:

1) Find the instruction in the listing which contains the record number to be changed.
2) Punch in an alter card (ALTER in columns 8 - 12) in columns 16 on the alter number of the instruction. This is found to the left of the symbolic address field in the listing.
3) Repunch the card exactly as before, replacing the old number with the new.
4) Insert the two cards in the mod package before running.
5) Ex. To change the record number of the fifth selection of file 4:

```
Column 8        16
       ALTER    1511,1511
       TXL      SKIP 1,4,"X"
```

where "X" is the new number.

Rome Air Development Center, Griffiss AF Base, NY
Rpt No. RADC-TDR-62-567, OPTIMUM SPEECH SIGNAL MAPPING TECHNIQUES, Final Technical Report, 28 Sep 62, 112pp incl illus
Unclassified Report

Investigations of the analysis of speech in terms of a fixed exponential function series have been carried out. The analysis-synthesis processing was performed via digital computers operating with digitized speech. The results indicate that the representation used is an efficient one for speech waveform analysis and that the information content of the speech is preserved when phase information is eliminated. The spectral coefficients after phase elimination are not found to be an efficient representation for the amplitude spectrum. Other results show that the method of analysis is not limited to the speech of one individual. Analytical studies indicate that it is possible to optimize the method of analysis to essentially perfect its efficiency for speech waveform analysis. Normalized, phase eliminated, spectral patterns derived for ten vowel utterances by five talkers indicate the feasibility of performing both automatic vowel and/or automatic speaker recognition using the orthonormal coefficient data.

1. Speech Representation
2. Acoustic Filters
I. Project 4027
Task 402704
II. Cont AF30(602)-2715
III. Sylvania Electronic Systems, Waltham 54, Mass.
IV. Secondary Rpt No. F-1005-1
V. In ASTIA collection

Rome Air Development Center, Griffiss AF Base, NY
Rpt No. RADC-TDR-62-567, OPTIMUM SPEECH SIGNAL MAPPING TECHNIQUES, Final Technical Report, 28 Sep 62, 112pp incl illus
Unclassified Report

Investigations of the analysis of speech in terms of a fixed exponential function series have been carried out. The analysis-synthesis processing was performed via digital computers operating with digitized speech. The results indicate that the representation used is an efficient one for speech waveform analysis and that the information content of the speech is preserved when phase information is eliminated. The spectral coefficients after phase elimination are not found to be an efficient representation for the amplitude spectrum. Other results show that the method of analysis is not limited to the speech of one individual. Analytical studies indicate that it is possible to optimize the method of analysis to essentially perfect its efficiency for speech waveform analysis. Normalized, phase eliminated, spectral patterns derived for ten vowel utterances by five talkers indicate the feasibility of performing both automatic vowel and/or automatic speaker recognition using the orthonormal coefficient data.

1. Speech Representation
2. Acoustic Filters
I. Project 4027
Task 402704
II. Cont AF30(602)-2715
III. Sylvania Electronic Systems, Waltham 54, Mass.
IV. Secondary Rpt No. F-1005-1
V. In ASTIA collection

Rome Air Development Center, Griffiss AF Base, NY
Rpt No. RADC-TDR-62-567, OPTIMUM SPEECH SIGNAL MAPPING TECHNIQUES, Final Technical Report, 28 Sep 62, 112pp incl illus
Unclassified Report

Investigations of the analysis of speech in terms of a fixed exponential function series have been carried out. The analysis-synthesis processing was performed via digital computers operating with digitized speech. The results indicate that the representation used is an efficient one for speech waveform analysis and that the information content of the speech is preserved when phase information is eliminated. The spectral coefficients after phase elimination are not found to be an efficient representation for the amplitude spectrum. Other results show that the method of analysis is not limited to the speech of one individual. Analytical studies indicate that it is possible to optimize the method of analysis to essentially perfect its efficiency for speech waveform analysis. Normalized, phase eliminated, spectral patterns derived for ten vowel utterances by five talkers indicate the feasibility of performing both automatic vowel and/or automatic speaker recognition using the orthonormal coefficient data.

1. Speech Representation
2. Acoustic Filters
I. Project 4027
Task 402704
II. Cont AF30(602)-2715
III. Sylvania Electronic Systems, Waltham 54, Mass.
IV. Secondary Rpt No. F-1005-1
V. In ASTIA collection

Rome Air Development Center, Griffiss AF Base, NY
Rpt No. RADC-TDR-62-567, OPTIMUM SPEECH SIGNAL MAPPING TECHNIQUES, Final Technical Report, 28 Sep 62, 112pp incl illus
Unclassified Report

Investigations of the analysis of speech in terms of a fixed exponential function series have been carried out. The analysis-synthesis processing was performed via digital computers operating with digitized speech. The results indicate that the representation used is an efficient one for speech waveform analysis and that the information content of the speech is preserved when phase information is eliminated. The spectral coefficients after phase elimination are not found to be an efficient representation for the amplitude spectrum. Other results show that the method of analysis is not limited to the speech of one individual. Analytical studies indicate that it is possible to optimize the method of analysis to essentially perfect its efficiency for speech waveform analysis. Normalized, phase eliminated, spectral patterns derived for ten vowel utterances by five talkers indicate the feasibility of performing both automatic vowel and/or automatic speaker recognition using the orthonormal coefficient data.

1. Speech Representation
2. Acoustic Filters
I. Project 4027
Task 402704
II. Cont AF30(602)-2715
III. Sylvania Electronic Systems, Waltham 54, Mass.
IV. Secondary Rpt No. F-1005-1
V. In ASTIA collection